ETALON

D3.1 Trade-off Analysis for On-board and Track-side Communication Systems

Due date of deliverable: 28/02/2018

Actual submission date: 26/02/2018

Leader of this Deliverable: ISMB

Reviewed: Yes

Document status		
Revision	Date	Description
1	14.11.17	First ToC shared within the consortium.
2	17.11.17	Integration of the contributions relative to the On-board and Track-side communication by UNEW and ARD.
3	11.12.17	Integration of the contributions relative to the On-board communication by PER and ISMB.
4	13.12.17	First revision of the document.
5	31.01.18	Integrated final contributions, added the conclusions chapter and sent for internal review.
6	16.02.18	Addressed comments from internal review, cleaned the document from the remaining comments and finalised some pending entries.
7	23.02.18	Final internal check before the submission.
8	26.02.18	Quality Check

Project funded from the European Union’s Horizon 2020 research and innovation programme

Dissemination Level

PU	Public	X
CO	Confidential, restricted under conditions set out in Model Grant Agreement	
CI	Classified, information as referred to in Commission Decision 2001/844/EC	

Start date of project: 01/09/2017 Duration: 24 months

REPORT CONTRIBUTORS

Name	Company	Details of Contribution
Alexander James Pane	ISMB	Document generation, added the state of the art contribution related to the on-board communication systems and reviewed the track-side communication system contributions.
Paul Hyde	UNEW	Added the current market solutions related to the on-board communication systems and reviewed the existing content.
David Vincent	PER	Added some contributions related to the on-board communication systems and reviewed the existing content. Corrected some typing errors and added some specific content regarding the Perpetuum “String” solution.
Veronika Nedviga	ARD	Added the content for the track-side communication systems and reviewed the existing content. Updated the content for the track-side communication systems.
Carles Artigas	ARD	Added comments and content regarding secure communication over WSN.
Roberto Cafferata	SIRTI	Added general comments and content to the document
Francesco Sottile	ISMB	Performed a final check of the whole document.
Klevisa Ceka	RINA – C BE	Performed the Quality Check of the whole

		document.

EXECUTIVE SUMMARY

The main objective of the ETALON WP3 focuses on the investigation and development methods to check the integrity of a train, with a final aim of design, simulate and prototype a wireless communication platform for information exchange on and off the train. This document represents the output of the Task 3.1, where the work is focused on the analysis of wireless communication solutions offered both from the market and the research community.

The purpose of this document is to examine the offered solutions and to focus on the most relevant ones for the ETALON project. The document firstly introduces the basic concepts of wireless technology. Afterwards, two analyses are performed for both On-board and Track-side wireless communication solutions.

The trade-off analysis takes under consideration many factors, such as energy consumption, reliability of the network and employed technology for both hardware and software. Each solution is then categorized based on the possibility of applying it to the ETALON project prospective. It is worth remarking that the state of the art solutions do not employ energy harvesting technologies, while this is a key aspect of the ETALON project. Furthermore, safety aspects have not taken into account in the analysed solutions, while the ETALON system must be SIL4 “able”.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF PARTICIPANTS

NO	LEGAL NAME	SHORT NAME
2	Sirti Società per Azioni	SIRTI
3	Ardanuy Ingineria SA	ARD
6	Istituto Superiore Mario Boella Sulle Tecnologie Dell'Informazione e delle Telecomunicazioni Associazione	ISMB
7	Perpetuum Limited	PER
8	University of Newcastle upon Tyne	UNEW

List of Acronyms

Acronym	Meaning
ACK	Acknowledgment packet
ACL	Access List
AES	Advanced Encryption Standard
AREA-MAC	Asynchronous, Real-time, Energy-efficient and Adaptive Medium Access Control
ATP	Automatic Train Protection
BSN	Basic Sensor Nodes
CA	Consortium Agreement
CBC	Cipher Block Chaining
CCM	Counter with CBC-MAC
CDF	Cumulative Density Function
CDMA	Code Division Multiplexing Access
CEN	European Committee for Standardisation

CTR	Counter mode
CTS	Clear To Send
DDN	Data Dissemination Nodes
DoS	Denial of Service
DRN	Data Relay Nodes
EC	European Commission
EC-MQV	Elliptic Curve Menezes-Qu-Vanstone
GA	Grant Agreement
GPS	Global Positioning System
IT	Interarrival Time
IXL	Interlocking
LPL	Low-Power Listening
LSN/LWSN	Linear Wireless Sensor Network
MAC	Medium Access Control
MCS	Modulation and Coding Scheme
MiTM	Man in the middle
OSI	Open Systems Interconnection
PA	Power Amplifier
PDR	Packet Delivery Ratio
PMO	Project Management Office
QM	Quality Manager
RB	Resource Block
RF	Radio Frequency
RSS	Received Signal Strength
RSSI	Received Signal Strength Indicator

RTS	Request To Send
SC	Steering Committee
S-MAC	Sensor Medium Access Control
SNR	Signal to Noise Ratio
SPE	Segment Primary Node
SSE	Segment Secondary Node
SWOC	Smart Wayside Object Controller
TBS	Transport Block Size
TD_ACK	Topology Discovery Acknowledgment message
TD_HELLO	Topology Discovery Hello message
TD_HELLO_COUNT	Topology Discovery Count message
TD_L1_COMP	Topology Discovery L1 Completion message
TDMA	Time Division Multiplexing Access
TMS	Train Management System
TMT	Technical Management Team
WP	Work Package
WPL	Work Package Leader
WSN	Wireless Sensor Network

1. INTRODUCTION

The document is organised as follows. Section 2 introduces the basic concepts of wireless technologies and outlines the existing protocols and solutions.

Section 3 focuses on the study for On-board Wireless Communication solutions. This section includes sub-section 3.2 that analyses already existing solutions and subsection 3.3 that analyses research solutions. Finally, sub-section 3.4 presents a trade-off analysis outlining the applicability of the analysed solutions to the ETALON purpose.

Section 4 focuses on the study for Track-side Wireless Communication solutions. Following the same structure of the On-board analysis, this section is also composed by two sub-sections. In particular, sub-section 4.2 analyses already existing solutions and subsection 4.3 analyses research solutions. Finally, a trade-off analysis is presented in subsection 4.4, outlining the applicability of the analysed solutions to the ETALON purpose.

Finally, section 5 draws the conclusions of the analysis, outlining all the strong and weak points of the presented solutions.

2. ANALYSIS OF RELEVANT WIRELESS TECHNOLOGIES

Various wireless technologies that can be employed for the Radio Frequency (RF) module are available for the ETALON project. A first analysis of these technologies and possible limitations and applications are investigated in order to better understand the best fitting RF technology that can be employed for the ETALON project.

2.1 THE OSI MODEL

Various protocols are available for the creation of a wireless network, an analysis can be performed using the OSI model. The model presents two main layers that contain sub-layers. The model is summarized in Table 1.

The two main layers are the Host and Media Layers. The first one is itself composed by another four layers, while the second one is composed by another 3 layers.

Table 1: OSI Model

Layer		Protocol data unit (PDU)	Function
Host layers	Application	Data	High-level APIs, including resource sharing, remote file access
	Presentation		Translation of data between a networking service and an application; including character encoding, data compression and encryption/decryption
	Session		Managing communication sessions, i.e. continuous exchange of information in the form of multiple back-and-forth transmissions between two nodes
	Transport	Segment (TCP) / Datagram (UDP)	Reliable transmission of data segments between points on a network, including segmentation, acknowledgement and multiplexing
Media layers	Network	Packet	Structuring and managing a multi-node network, including addressing, routing and traffic control
	Data link	Frame	Reliable transmission of data frames between two nodes connected by a physical layer
	Physical	Bit	Transmission and reception of raw bit streams over a physical medium

2.1.1 Layer 1: Physical Layer

The physical layer defines the electrical and physical specifications of the data connection. It defines the relationship between a device and a physical transmission medium (for example, an electrical cable, an optical fibre cable, or a radio frequency link). This includes the layout of pins, voltages, line impedance, cable specifications, signal timing and similar characteristics for connected devices and frequency for wireless devices. It is responsible of the transmission and reception of unstructured raw data in a physical medium. Bit rate control is performed at the physical layer. It may define transmission mode as simplex, half duplex, and full duplex, it also defines the network topology.

2.1.2 Layer 2: Data Link Layer

The data link layer provides node-to-node data transfer, a link between two directly connected nodes. It detects and possibly corrects errors that may occur in the physical layer. It defines the protocol to establish and terminate a connection between two physically connected devices. It also defines the protocol for flow control between them.

IEEE 802 divides the data link layer into various sub-layers, where the mostly known are:

- Medium access control (MAC) layer – responsible for controlling how devices in a network gain access to a medium and permission to transmit data.
- Logical link control (LLC) layer – responsible for identifying and encapsulating network layer protocols, and controls error checking and frame synchronization.

2.1.3 Layer 3: Network Layer

The network layer provides the functional and procedural means of transferring variable length data sequences (called datagrams) from one node to another connected in "different networks". A network is a medium to which many nodes can be connected, on which every node has an address, which permits nodes connected to it to transfer messages to other nodes connected to it by merely providing the content of a message and the address of the destination node and letting the network find the way to deliver the message to the destination node, possibly routing it through intermediate nodes. If the message is too large to be transmitted from one node to another on the data link layer between those nodes, the network may implement message delivery by splitting the message into several fragments at one node, sending the fragments independently, and reassembling the fragments at another node. It may, but does not need to, report delivery errors.

Message delivery at the network layer is not necessarily guaranteed to be reliable; a network layer protocol may provide reliable message delivery, but it does not need to do so.

2.1.4 Layer 4: Transport Layer

The transport layer provides the functional and procedural means of transferring variable-length data sequences from a source to a destination host via one or more networks, while maintaining the quality of service functions.

The transport layer controls the reliability of a given link through flow control, segmentation/de-segmentation, and error control. Some protocols are state-oriented and connection-oriented. This means that the transport layer can keep track of the segments and re-transmit those that fail. The transport layer also provides the acknowledgement of the successful data transmission and sends the next data if no errors occurred. The transport layer creates packets out of the message received from the application layer. Packetizing is a process of dividing the long message into smaller messages.

The OSI defines five classes of connection-mode transport protocols. Class 0 contains no error recovery and was designed for use on network layers that provide error-free connections, while class 4 is closest to TCP, although TCP contains functions, such as the graceful close, which OSI assigns to the session layer (layer 5). A detailed table can be viewed in Table 2.

Table 2: Transport classes for Layer 4

	Recovery from disconnection and reset	Multiplex several transport onto one network connection	Error recovery
Class 0	NO	NO	NO
Class 1	YES	NO	NO
Class 2	NO	YES	NO
Class 3	YES	YES	NO
Class 4	YES	YES	YES

2.1.5 Layer 5: Session Layer

The session layer controls the dialogues (connections) between computers. It establishes, manages and terminates the connections between the local and remote application. It provides for full-duplex, half-duplex, or simplex operation, and establishes checkpointing, adjournment, termination, and restart procedures. The OSI model made this layer responsible for the graceful closure of sessions, which is a property of the Transmission Control Protocol, and for session checkpointing and recovery, which is not usually used in the Internet Protocol Suite. The session layer is commonly implemented explicitly in application environments that use remote procedure calls.

2.1.6 Layer 6: Presentation Layer

The presentation layer establishes context between application-layer entities, in which the application-layer entities may use different syntax and semantics if the presentation service provides a mapping between them. If a mapping is available, presentation service data units are encapsulated into session protocol data units and passed down the protocol stack.

This layer provides independence from data representation by translating between application and network formats. The presentation layer transforms data into the form that the application accepts. This layer formats data to be sent across a network. It is sometimes called the syntax layer. The presentation layer can include compression functions. The Presentation Layer negotiates the Transfer Syntax.

2.1.7 Layer 7: Application Layer

The application layer is the OSI layer closest to the end user, which means both the OSI application layer and the user interact directly with the software application. This layer interacts with software applications that implement a communicating component. Such application programs fall outside the scope of the OSI model. Application-layer functions typically include identifying communication partners, determining resource availability, and synchronizing communication. When identifying communication partners, the application layer determines the identity and availability of communication partners for an application with data to transmit.

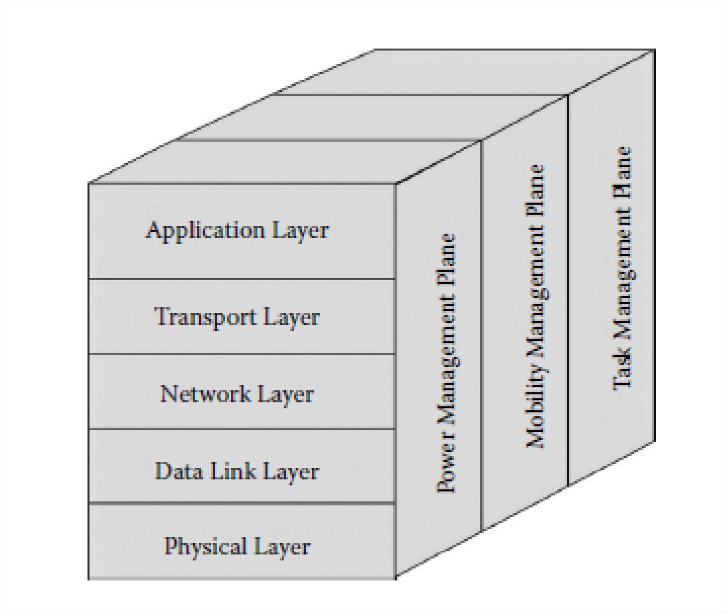


Figure 1: Sensor Network Protocol Stack.

2.2 Analysis of the Wanted OSI Features

From the previously presented model, an analysis of the main wanted features can be performed. To implement the protocol not all layers are required, so just the wanted feature can be implemented, developing an ad-hoc protocol.

Starting from the lowest two layers (the Physical and Data Link Layers), for the ETALON project a wireless communication is physically required, thus reducing the possible choices to the following:

- IEEE 802.11

- IEEE 802.15

2.2.1 The IEEE 802.11 Standard

The IEEE 802.11 is a set of MAC (Media Access Control) in Layer 2 and employs in Layer 1 a wireless communication for implementing a WLAN (Wireless Local Area Network). The IEEE 802.11 works in five frequency ranges:

- 900 MHz (IEEE 802.11ah)
- 2.4 GHz (IEEE 802.11b, IEEE 802.11g and IEEE 802.11n)
- 3.6 GHz (IEEE 802.11y-2008)
- 5 GHz (IEEE 802.11a and IEEE 802.11ac)
- 60 GHz (IEEE 802.11ad – Wireless Gigabit Alliance)

The IEEE 802.11 is widely used as the Wi-Fi standard for home routers. The family consists of a series of half-duplex over-the-air communication techniques that all use the basic 802.11 protocol.

This protocol has an approximate range of 20 – 70 meters indoor and 100 – 5000 meters outdoor, with a minimum data rate of 1 Mbit/s to even Gbit/s data rates. Such high-speed data rates do though come with a high energy consumption drawback, limiting the application of the IEEE 802.11 standard to environments where energy consumption is not critical and devices have direct access to a main power supply.

It must also be considered that the used frequencies are also employed for other typologies of usage, introducing also an interference problem with external devices, that belong to the railway, like Eurobalises [21], or are employed for other uses.

Additionally, the main used frequencies (the 2.4 GHz and 5 GHz) are not suitable (short wavelengths so not diffract easily around metal bogie components) for the application in a harsh environment, such as the railway environment. This would limit the sensing range and limit the network that needs to be as stable as possible in any environment.

2.2.2 The IEEE 802.15 Standard

The IEEE 802.15 can be divided in sub-standards:

- IEEE 802.15.1 – WPAN / Bluetooth
- IEEE 802.15.2 – Coexistence
- IEEE 802.15.3 – High rate WPAN
- IEEE 802.15.4 – Low rate WPAN
- IEEE 802.15.5 – Mesh networking
- IEEE 802.15.6 – Body area networks
- IEEE 802.15.7 – Visible light communication
- IEEE 802.15.8 – Peer aware communication

- IEEE 802.15.9 – Key management protocol
- IEEE 802.15.10 – Layer 2 routing

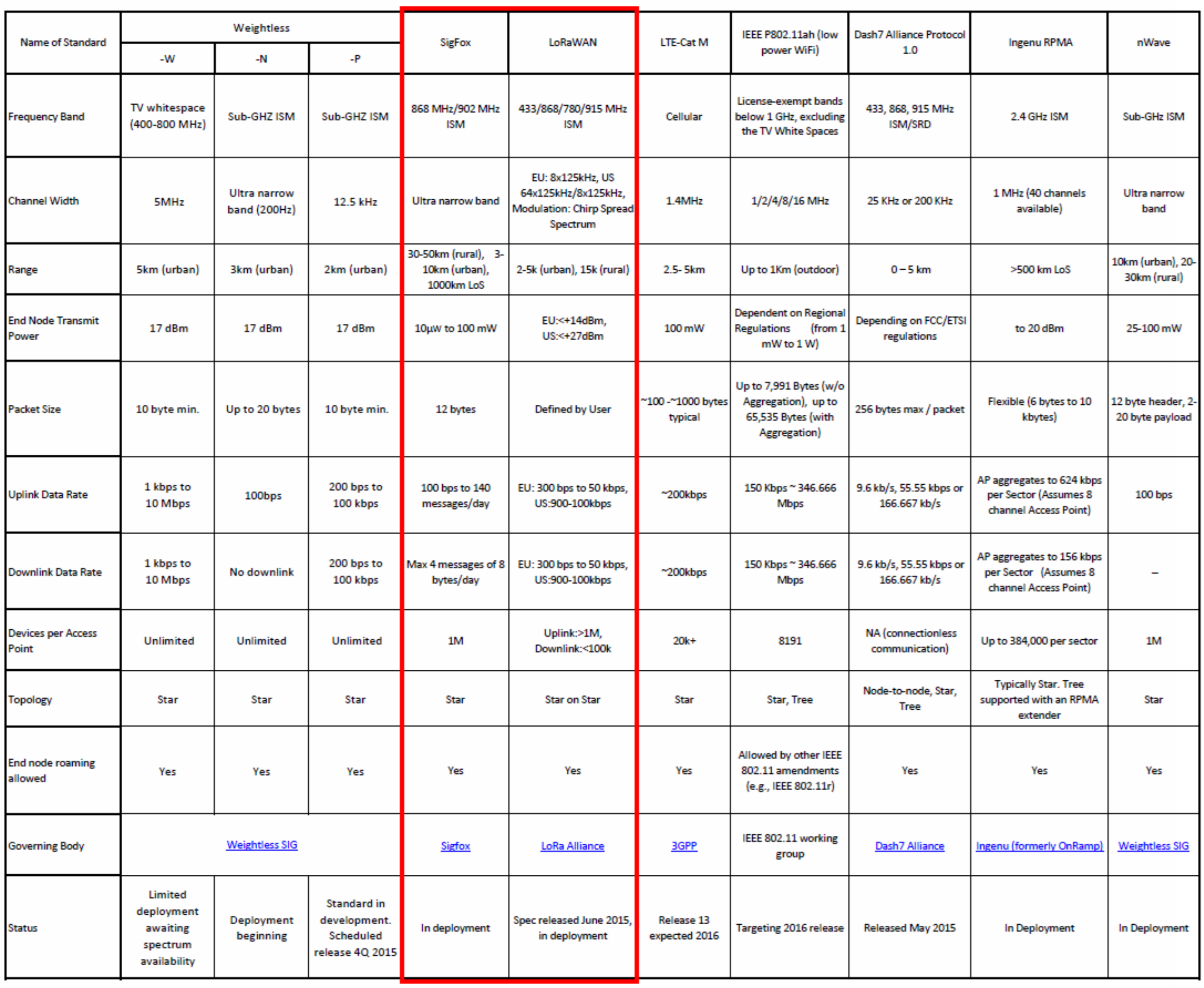
The listed standards defined various layers of the OSI model, majorly they all define the first two layers, the Physical and Data Link layers.

Regarding the employed frequencies, these vary from the usual 2.4 and 5 GHz bands, to the country dependent sub-GHz frequencies, which are 868 MHz in Europe and 915 MHz in the US [20].

2.3 Full OSI Protocols

Protocols that already implement multiple layers of the OSI model exist, such as ZigBee, 6LoWPAN, Z-Wave and XBee.

- ZigBee
- 6LoWPAN
- Z-Wave
- Xbee
- SigFox
- Dash7
- Thread
- Insteon
- Z-Wave
- Lorawan
- Enocean



Name of Standard	Weightless			SigFox	LoRaWAN	LTE-Cat M	IEEE P802.11ah (low power WiFi)	Dash7 Alliance Protocol 1.0	Ingenu RPMA	nWave
	-W	-N	-P							
Frequency Band	TV whitespace (400-800 MHz)	Sub-GHZ ISM	Sub-GHZ ISM	868 MHz/902 MHz ISM	433/868/780/915 MHz ISM	Cellular	License-exempt bands below 1 GHz, excluding the TV White Spaces	433, 868, 915 MHz ISM/SRD	2.4 GHz ISM	Sub-GHz ISM
Channel Width	5MHz	Ultra narrow band (200Hz)	12.5 kHz	Ultra narrow band	EU: 8x125kHz, US 64x125kHz/8x125kHz, Modulation: Chirp Spread Spectrum	1.4MHz	1/2/4/8/16 MHz	25 KHz or 200 KHz	1 MHz (40 channels available)	Ultra narrow band
Range	5km (urban)	3km (urban)	2km (urban)	30-50km (rural), 3-10km (urban), 1000km LoS	2-5k (urban), 15k (rural)	2.5- 5km	Up to 1Km (outdoor)	0 – 5 km	>500 km LoS	10km (urban), 20-30km (rural)
End Node Transmit Power	17 dBm	17 dBm	17 dBm	10μW to 100 mW	EU:<+14dBm, US:<+27dBm	100 mW	Dependent on Regional Regulations (from 1 mW to 1 W)	Depending on FCC/ETSI regulations	to 20 dBm	25-100 mW
Packet Size	10 byte min.	Up to 20 bytes	10 byte min.	12 bytes	Defined by User	~100 -~1000 bytes typical	Up to 7,991 Bytes (w/o Aggregation), up to 65,535 Bytes (with Aggregation)	256 bytes max / packet	Flexible (6 bytes to 10 kbytes)	12 byte header, 2-20 byte payload
Uplink Data Rate	1 kbps to 10 Mbps	100bps	200 bps to 100 kbps	100 bps to 140 messages/day	EU: 300 bps to 50 kbps, US:900-100kbps	~200kbps	150 Kbps ~ 346.666 Mbps	9.6 kb/s, 55.55 kbps or 166.667 kb/s	AP aggregates to 624 kbps per Sector (Assumes 8 channel Access Point)	100 bps
Downlink Data Rate	1 kbps to 10 Mbps	No downlink	200 bps to 100 kbps	Max 4 messages of 8 bytes/day	EU: 300 bps to 50 kbps, US:900-100kbps	~200kbps	150 Kbps ~ 346.666 Mbps	9.6 kb/s, 55.55 kbps or 166.667 kb/s	AP aggregates to 156 kbps per Sector (Assumes 8 channel Access Point)	–
Devices per Access Point	Unlimited	Unlimited	Unlimited	1M	Uplink:>1M, Downlink:<100k	20k+	8191	NA (connectionless communication)	Up to 384,000 per sector	1M
Topology	Star	Star	Star	Star	Star on Star	Star	Star, Tree	Node-to-node, Star, Tree	Typically Star. Tree supported with an RPMA extender	Star
End node roaming allowed	Yes	Yes	Yes	Yes	Yes	Yes	Allowed by other IEEE 802.11 amendments (e.g., IEEE 802.11r)	Yes	Yes	Yes
Governing Body	Weightless SIG			Sigfox	LoRa Alliance	3GPP	IEEE 802.11 working group	Dash7 Alliance	Ingenu (formerly OnRamp)	Weightless SIG
Status	Limited deployment awaiting spectrum availability	Deployment beginning	Standard in development. Scheduled release 4Q 2015	In deployment	Spec released June 2015, in deployment	Release 13 expected 2016	Targeting 2016 release	Released May 2015	In Deployment	In Deployment

2.4 The Network Topology

The network topology used for the on-board communication system must be generated while the train is at a standstill. For the topology creation the problem of identifying if a node is part of the network arises, since in this scenario there could be various trains at a standstill and each of them has their own devices mounted. Also, the possibility where near trains may start the network topology creation at the same time could arise, but each network must be able to distinguish the nodes that belong to their own network.

One solution as exposed from the previous research activities, is to analyse the train acceleration using accelerometers and finding the correlation. This though has a main drawback: if trains are at a standstill all their accelerations are null, so it is not possible to distinguish the affinity of each node. A solution could be to generate the network topology when the train starts moving, so that the correlation of the acceleration is able to distinguish the network nodes. Although, the device will be located on the train bogies, where a lot of vibrations are present and so finding the corresponding train acceleration can be very challenging. Additionally, adding accelerometers and the required level of signal processing would add significantly to the cost and complexity of the devices when simpler solutions (such a pre-defined vehicle/train topology) might be possible.

Another solution is to know the ID or address of each node during the installation of the device, so that each node will be provided with a list of devices that belong to the network, thus neglecting all the devices in range that do not belong to the provided list. This list can be generated by either inserting the values directly in each device (using RFID or similar short-range communication) or by inserting the whole list in the coordinator node (on the locomotive) so that this node can start the communication with the other devices in the list and pass along the list. The initial data containing the list of device identifications be passed to the coordinator also using RFID or other short-range communication. In this scenario the installation process will have a device registration phase, where the installer will register all nodes (using an ad-hoc registration device) apart the coordinator one. Once all the data is stored in the registration device, this can be connected to the coordinator node and all the information gathered from the nodes can be passed. At this point, the coordinator has the knowledge of the identity of all the nodes, so it can start the process of network topology generation.

On one side, this solution adds a small overhead for the installation process, but this results in a quite simple procedure. On the other side, this procedure prevents beforehand the possibility of unwanted communication with nodes that don’t belong to the network, so that the network topology can be correctly structured.

The level of manual interaction required for this should be considered, since train integrity is a safety critical operation. Human error must be fail safe, if possible in the process.

2.5 Long Term Evolution - Advanced (LTE-A)

LTE-A is an evolution of 3GPP-LTE which aims to bridge the gap between Third Generation (3G) and Fourth Generation (4G) standards. LTE-A aims to provide peak data rates of up to 1 Gbps (for low mobility) and 100 Mbit/s (for high mobility) in Downlink (DL) and 500 Mbps in Uplink (UL). LTE-A is required to reduce the latency time as compared to 3GPP-LTE. LTE-Advanced targets to enhance the cell edge user throughput in order to achieve a homogeneous user experience in the cell and increase the capacity to 30 and 15 bps/Hz in DL and UL, respectively.

The main new functionalities introduced by LTE-A to enhance 3GPP-LTE are carrier aggregation (CA), enhanced use of multi-input multi-output (MIMO) antenna techniques, coordinated multipoint transmission and reception (CoMP), and support by relay technology

- Carrier Aggregation: In carrier aggregation, multiple carrier components are aggregated, to provide wider bandwidths for transmission purposes both in DL and UL.
- Extended MIMO: LTE-A introduced extending the number of layers in MIMO from 4x4 to 8x8 layer at DL and from 2x2 to 4x4 layers at UL to increase the overall bit rate through transmission of different data streams in multiple antennas
- Coordinated Multipoint Transmission/Reception (CoMP): In CoMP multiple geographically separated base station sites coordinate transmission and reception, in order to achieve good system performance and end-user service quality.

- Relaying Technology: This technique provides the possibility for heterogeneous network planning through the integration of large cells such as BS, and small cells such as Relay Node (RN). RNs have lower power and a lower cost compared to BSs which provide enhanced coverage and capacity in cellular networks.

The design of the LTE physical layer (PHY) is heavily influenced by the requirements for high peak transmission rate, spectral efficiency, and multiple channel bandwidths. To fulfill these requirements, orthogonal frequency division multiplex (OFDM) was selected as the basis for the PHY layer. Together, OFDM and MIMO are two key technologies featured in LTE and constitute major differentiation over 3G systems which are based on code division multiple access (CDMA).

There are different modes of operation (FDD/TDD) and different downlink and uplink access technologies (OFDMA, SC-FDMA), along with options and exceptions for each mode and access technology.

2.5.1 Multiple Access Techniques

The OFDM technology is based on using multiple narrow band sub-carriers spread over a wide channel bandwidth. The sub-carriers are mutually orthogonal in the frequency domain which mitigates inter-symbol interference (ISI) as shown in Figure 2. The downlink physical layer of LTE is based on OFDMA and the uplink is based in single carrier frequency division access (SC-FDMA). The bandwidth of the single carrier is determined based on the required data rate by the user. This leads to similar link performance parameters for the uplink and the downlink.

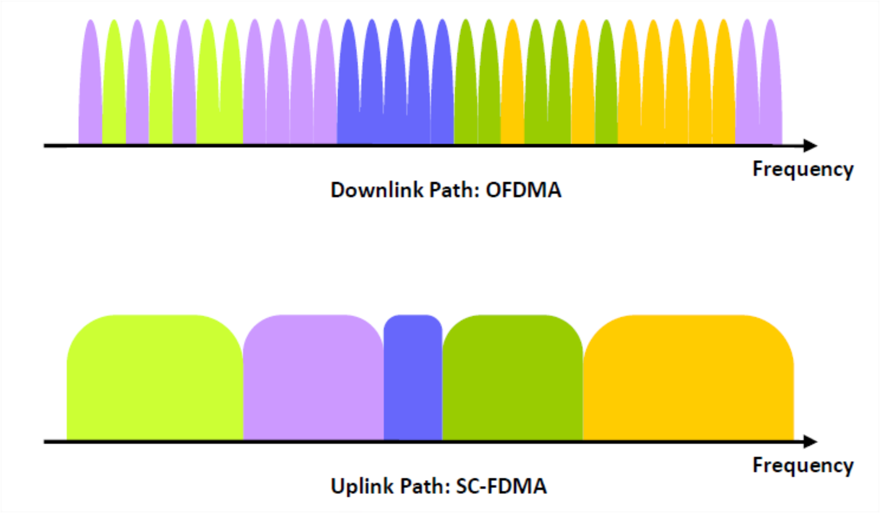


Figure 2: OFDM Subcarrier spacing.

The transmission can be scheduled by Resource Blocks (RB) each of which consists of 12 consecutive sub-carriers, or 180 kHz, for the duration of one slot (0.5 ms).

The smallest unit of resource is the Resource Element which consists of one SC-FDMA data block length on one sub-carrier. resource block consists of 12 REs for the duration of a slot (0.5 ms). In the time domain, a 10 ms uplink frame consists of 10 one ms subframes and 20 slots.

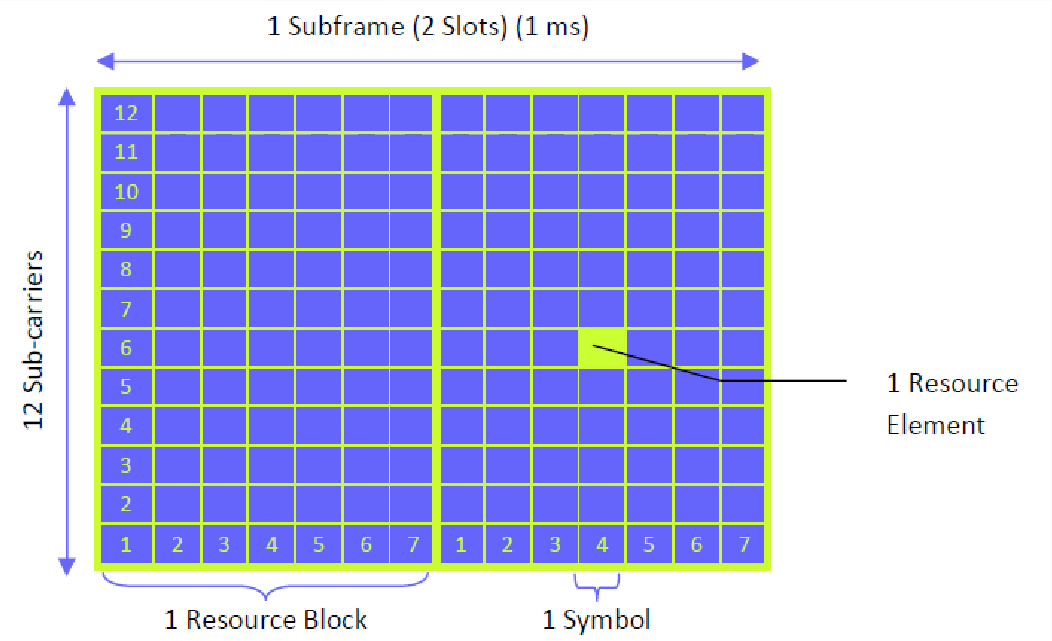


Figure 3: Subframe of LTE.

3. TRADE-OFF ANALYSIS FOR ON-BOARD COMMUNICATION SYSTEMS

3.1 INTRODUCTION

The new requirements introduced with the ETCS L3 standard have been a subject of interest for various research institutes and universities. At this point, a State of the Art has been conceptualised, summarising the research developed up to this point for resolving the issues introduces with ETCS L3.

3.2 CURRENT SOLUTIONS

3.2.1 Passenger Multiple Unit and Passenger Trains with Trailing Coaches

Modern passenger multiple units have a train data bus which runs along the complete length of the train for driving commands, communications, other train functions and general train condition monitoring. This generally takes the form of a hard-wired link with jumper/connecting cables between vehicles, it is a fairly simple matter to include functions for communicating and detecting train integrity into the train data bus. On older units and locomotive hauled passenger trains, even if there is no data bus, there are wired connections between the vehicles for power and other functions which would only be a slightly more complicated matter to include functions for communicating and detecting train integrity. Since in both cases there is a wired connection, using this connection or adding/modifying it for train integrity functions does not add an operational restriction since the cables have to be connected anyway. For freight vehicles there generally isn't any sort of power or data cable, so adding one which would require connection would add significant expense, impose additional operational requirements to connect the cables and add a vulnerability to the reliability of the wagons due to the added potential for mechanical failure of the cables and connectors. Therefore, a wired connection along the length of a train is not the most suitable option for on-board communication on freight trains.

3.2.2 European Freight Train

For most freight trains operating in Europe the only means of communication throughout the length of the train is via the continuous air brake pipe, which runs along the length of each wagon and is connected between each wagon, with a closed isolating valve at each end. The air pressure in this pipe is used to supply the air brake systems in each wagon, and variation of the air pressure is used to apply the brakes on each wagon. The air pressure is normally at full operational pressure and a reduction in pressure applies the brakes, in the event of loss of train integrity the air pipe connection part is disconnected, the air pressure is vented to atmosphere automatically applying the brakes. The sudden loss of air pressure indicates a fault to the driver, one potential cause would be a loss of train integrity, and since this is the worst-case scenario it must be assumed to be the case unless proved otherwise. This system acts as both a failsafe mechanism and detection system for loss of train integrity, also as a crude form of communication of a fault. The system relies on all the air pipes being connected and only the isolation valves at each of them being closed, the ones between each vehicle must be open for the pipe to be continuous throughout the train and the system to be effective. For integration with ETCS Level 3, if the length of the train is known then full pressure in the train brake pipe at the locomotive would indicate that the train is complete and train integrity is confirmed which could be communicated to the traffic control system by the locomotive. However, this would have to take into account intentional variations in air pressure due to brake applications, also if the air pressure was intentionally fully discharged from the pipe, then there would be no confirmation of train integrity.

In addition, there is potential for the train integrity failsafe to be ineffective if the manual procedures to ensure all the pipes are connected and the isolation valves throughout the train are open. If an intermediate isolation valve is closed, then the pipe is not continuous and only the front portion will have operational brakes and a loss of integrity in the rear portion would not be detected. In this case if the procedures to detect this are not carried out correctly then a train might run without any form of loss or integrity detection on the rear portion of the train.

The European railway is generally opposed in employing the train’s brake pipe for anything apart controlling the brakes.

3.2.3 United States of America Freight Trains

An air pressurised failsafe train brake pipe like the one used in Europe described previously is also used on US Railroads, however an additional device is sometimes used called a “Smart End of Train” device. These devices are attached to the tail end of the train and connected to the air brake pipe, using the available connection at the uncoupled end of the last vehicle, so that the device is pressurised. The device has air pressure sensors and valves as well as two-way radio communication with a device in the driver compartment connected to their brake control equipment. Depending on the feature level of a particular end of train, the device can perform the following functions:

- Display the air pressure at the tail end of the train locally and to the driver.
- Confirm train integrity (indicate complete loss of air pressure)
- Communicate the speed of the last wagon (for comparison with the locomotive)
- Measure the length and change in length of train
- Tracking of the tail of the train via GSM network
- Release the air from the train brake pipe:
    - Fully to assist in making an emergency application of the brakes by fully venting the air (venting air from both the drivers end and tail end ensures the pressure drop propagates along the pipe quicker leading to a faster application).
    - Partially to make controlled brake applications.

The devices are transportable (can be carried by one person) and battery powered, some devices have an additional air turbine which bleads a small amount of air (not enough to apply the brakes) from the pipe to run a small generator to provide additional power to support its functions. Whilst the current “Smart End of Train” device in use on US Railroad have features and functions which are relevant to ETCS Level 3, the devices (and locomotive installations) would need to be made compatible with ERTMS. Also, these devices have been developed for very different railway operational conditions, where long infrequent trains are prevalent and there is less emphasis on mixed traffic routes. Therefore, it is questionable how effective they would be a part of a traffic management system intended for increasing capacity on European routes.

It should be noted that communication between the EOT and locomotive is via high power long wavelength radio, which may not be feasible on busy European rail.

3.2.4 Amsted Rail IONX® Asset Monitoring

The IONX® Edge T-Series remote asset monitoring system from Amsted Rail uses remote wireless sensor systems attached to rail freight vehicles which communicate with each other locally within the train. The network of sensors sends sensor information to a master unit on the locomotive which has GPS and cellular communication functionality to allow the sensors to be tracked and the parameters monitored remotely. The battery powered wireless sensor nodes are marketed as being easy-to-install, durable and compact, they are claimed to last over a decade and require no maintenance. It should be noted that whilst the system does provide a wireless communication network on a train, it is intended for condition monitoring and logistics tracking which are not safety critical functions in terms of train integrity and traffic management. Although there are some safety related functions in terms of the condition of hazardous freight and the monitoring of vehicle condition which affect the safety of the train, these parameters do not have to be reported as often or as reliably as train integrity. Also, as a condition monitoring system, the route of data from different trains is not significant, it is not an issue if data from one train is received by another and forwarded to the monitoring centre by that other locomotive. However, for a train integrity functionality, managing communications which relate to the integrity (or otherwise) of specific trains is critical.

3.3 Research Solutions

3.3.1 Trainspotting Solution

The University of Twenty, Netherlands, presented a paper in 2009 [3] that analyses and implements a WSN for trains. The approach allocates sensors on all wagons and the train integrity is checked by measuring synchronicity of movement of the different carriages. The WSN differs from usual implementations because of the linear topology of trains.

The train communication features some specific characteristics for the protocol, some nodes cover special roles. The only special role, called Coordinator, is covered by the locomotive, whose node initiates actions. Initially multiple Coordinators can be in the network simultaneously.

The network routing is based on direct neighbour communication or communication towards the locomotive, meaning that the routing algorithms result simpler. Packets will be transferred towards or away the locomotive.

Since the train presents a linear characteristic, packets will barely travel through different paths, thus meaning that the occurrence of multi-paths is seldom. The whole network doesn’t need to acknowledge the train topology, but it is required that just the locomotive has this information.

It is possible to combine synchronisation with other kinds of communication, thus reducing the overhead due to synchronisation.

Nodes don’t initialise communication but are requested information and move the request along the network; this simplifies the collision issue of the network.

Network traffic doesn’t depend on the environment events, traffic patterns are known beforehand.

In Figure 4 the used protocol for communication is outlined. In the first two steps the initialization phase is represented; first of all, the sensor data correlation and afterwards the discovery, by using a topology request from the locomotive.

In the remaining two steps it is possible to view how an operation request is processed by the network. The request is initialised from the locomotive towards the adjacent neighbour, the latter one passes the request to the next neighbour up to the last wagon. When the last wagon receives the request, then the message travel the same path but in opposite direction.

The initialisation phase is performed by broadcasting so that each sensor node becomes aware of its neighbours. During the discovery the request from the locomotive floods in the network where the hop counts are calculated.

The up-time for the radio must be limited to minimize power consumption, so the radio must be active only during data transmission and resending packets must be minimized. The nodes must then provide low-power and stand-by modes, so that the network can take advantage from these features.

Sophisticated MAC layer protocols deal with this kind of problem, such as S-MAC, L-MAC and D-MAC. The D-MAC is the most appropriate implementation for this purpose, the protocol provides an efficient path without sleep delays spaced by long sleep periods.

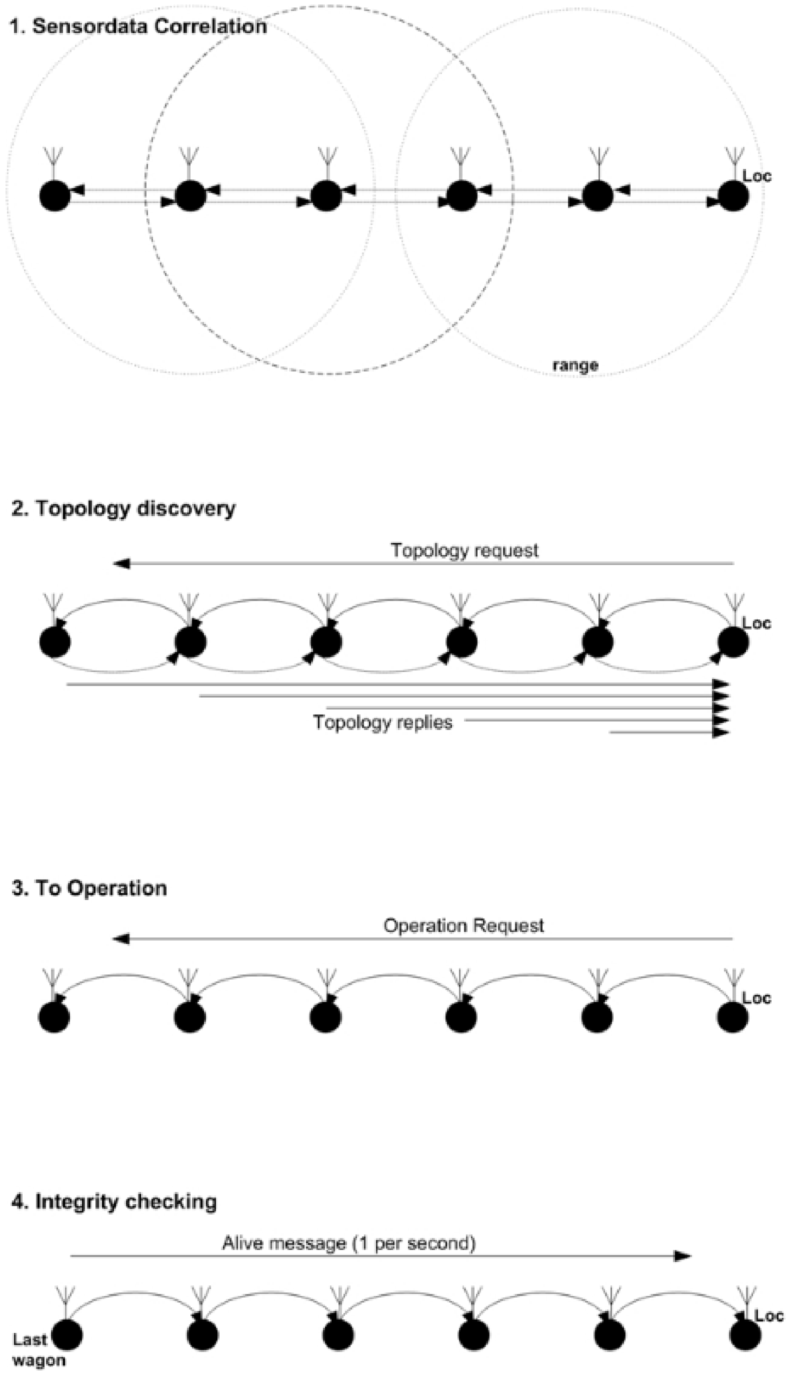


Figure 4: Communication Architecture.

Trade-off Analysis for the Trainspotting Solution

The solution proposed by University of Twenty is a good base for the generation of a WSN for Train Integrity, even though some issues are not taken much into account. The research takes into account only a single train in the sensing range, this is not though a real scenario. In a real scenario one or more trains can be in the WSN sending range, and if all nodes are operating on the same system (which would be required for compatibility), there would be no way of distinguishing the nodes that should belong to a particular WSN.

This potential solution is based on the exchange of an “alive” packet between the nodes, and can quickly acknowledge the absence of a node and thus confirm that the train integrity is not confirmed.

Another issue that the research does not consider is lack of power supply where the node is located. The power consumption and the optimisation of the nodes are taken in consideration, but it is not specified if the application can be battery based or be supplied by energy harvesting devices.

The study also does not consider problems introduced by failing nodes and no redundancy method is applied to prevent such problems.

The work also exposes the possibility that multiple coordinators could be active at the same time. This behaviour must be avoided in the ETALON project, where only the locomotive must cover the role of coordinator.

An issue to consider is also the possibility of saving multiple communication paths or introducing redundant sensors, this gives the possibility of having a more resilient network in case of node failure or nodes down for power saving or other reasons.

In the implementation of the network, the coordinator has the role of requesting and consequently receive data from the sensors, this requires data to travel from the head of the train (the coordinator) down to the tail and then back from the tail to the head. An alternative is to send the message from the tail towards the head, avoiding going through the network twice, thus saving energy.

The University of Twenty didn’t consider in this work the possibility of multiple trains performing topology discovery. If multiple trains are in sensing range, it is not handled how a coordinator should distinguish if a sensor is actually part of the consist or if belongs to another train. Another issue is represented by the interference, a solution with multiple channels should be considered, so that the most appropriate channel with less interference can be used for communication, taking also into account that long trains may present a variable interference along the consist.

3.3.2 Sensing Train Integrity Solution

The University of Twenty continued the previously outlined research with a second paper published in the same year [4].

This study moves from the usage of a WSN for gathering data to an application for train integrity sensing. Firstly, all the problems regarding “train composition” and “states and transitions” are outlined. Secondly, after finding the best fitting solution for these problems two days of experimentation were carried out to check the architecture results.

As previously said the problem of “train composition” must be solved to correctly check the train integrity, this means that the network must acknowledge the topology of the train, so that it excludes nodes that are not part of the train but are in sensing range.

The possibility of sensing nodes that are external to the train that are part of an active network monitoring the integrity of another train is highly probable. This scenario is possible when the train is at a stop in the station, where other trains are close by, or whenever another train might be travelling in the opposite direction on another track, thus being in that moment in the sensing range of the other train, an example case is depicted in Figure 5. In the work by University of Twenty, two possible methods for resolving the issue of determining train composition were considered: “localisation” or “Dynamic group awareness”

Using the localisation solution involves using a GPS module inside the node or using the RSSI parameter to compute the distance between nodes. The usage of a GPS module is not feasible, since the module would require an amount of energy higher than the quantity of resource available. On the other side using the RSSI value to compute distances and then use triangulation, trilateration or multilateration also presents various issues. The RF strength is very inaccurate, and the range is inconsistent, usually resulting in errors above 50%, also the topology of the train prohibits accurate triangulation, trilateration or multilateration. So, a solution based on “localisation” is discarded.

The second solution is using “Dynamic group awareness”. This solution is based on the principle that all the nodes that are part of the same train are highly correlated. The node correlation is based on the acceleration values given from the accelerometers of the nodes. Nodes that are part of the same train will present a high acceleration correlation. This solution appeared more appropriate and it was the one which was implemented for examination in the work by University of Twenty.

To check train integrity the study acknowledges that the train can be in three possible states: “initialisation”, “idle” or “operation”, the state transition diagram can be viewed in Figure 6.

Whenever the train is standing still it is in the “idle” state. In this state the WSN can determine the all possible configurations of the train, but since there is no movement there is no “group awareness”, so the solution may include wagons of other trains that are standing still nearby. Only when the train starts moving this information become available and the WSN can correctly determine the wanted wagons. When this process starts the “initialisation” phase is entered. There is though the rare possibility that other trains that is in the WSN sensing area start moving at the same time, but the difference in the movements should distinguish the various trains.

Now that the train composition is known from the “initialisation” phase, the state transitions to the “operation” state. At this point the train composition is known and only the wagons that are part of the network need to be checked, this can be optimised by saving the corresponding network data.

The “operation” state could run the same algorithm as the “initialisation” state, but this would require running correlation algorithms that require energy that may not be available. At this point in the study two days of experimental testing were carried out intended to test the various cases to check the various implementations.

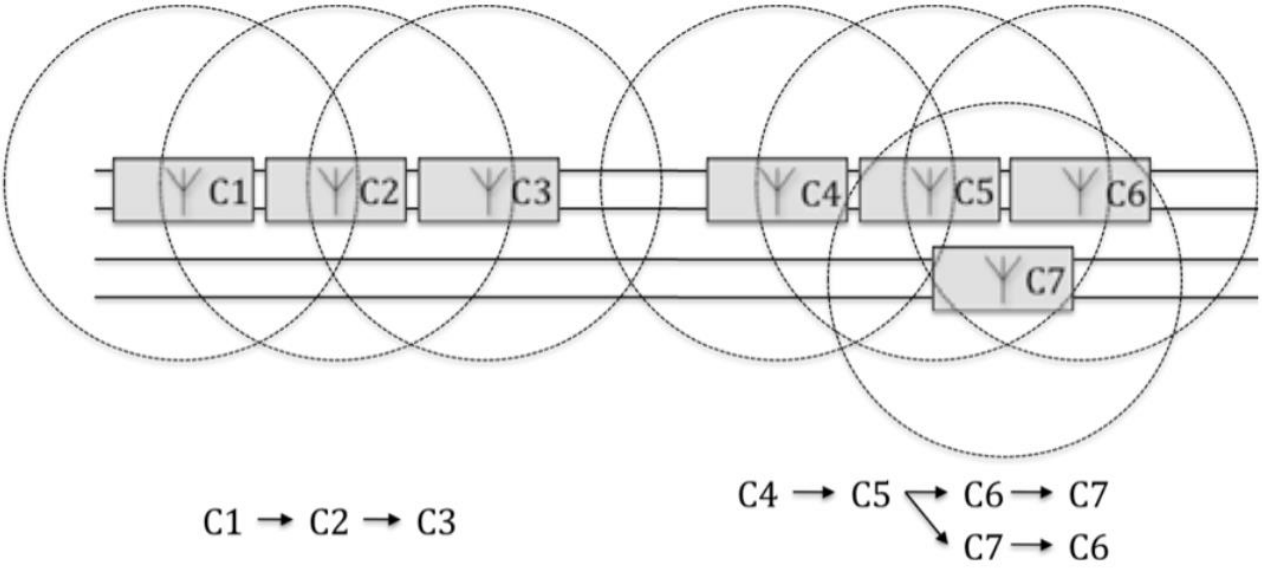


Figure 5: Possible train composition.

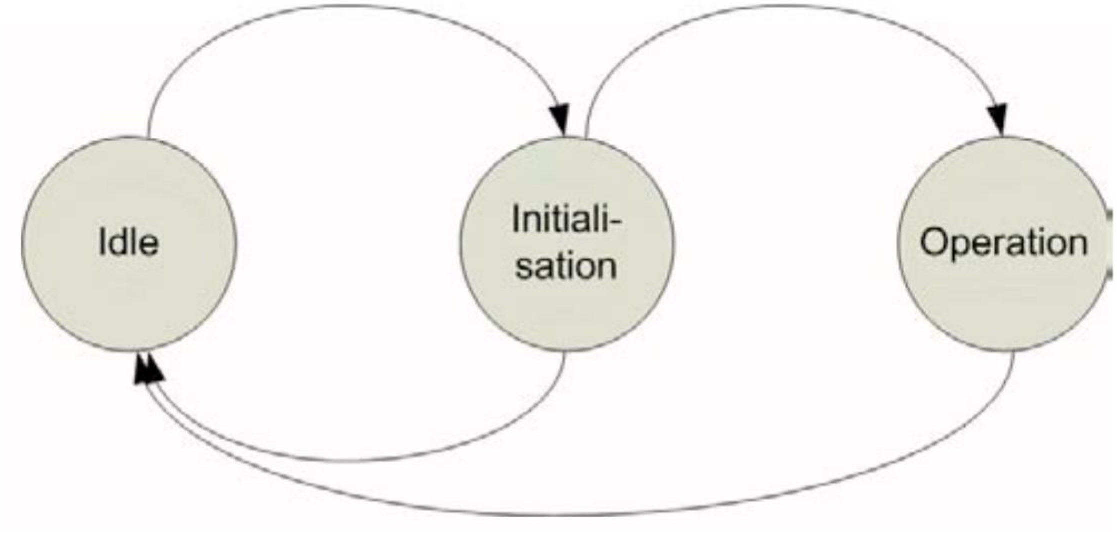


Figure 6: The three possible states to check train integrity.

Trade-off Analysis for the Sensing Train Integrity Solution

The solution proposed by University of Twenty considers the possibility of interference between WSN that belong to separate trains. The research proposed two solutions for this issue: localisation and dynamic group awareness. The localisation-based solution is discarded due to power issues, while the dynamic group awareness solution is applied.

This solution is based on the acceleration correlation, so that only devices that sensing the same moving direction are part of the network. This solution presents some issues, firstly in the scenario where other trains that are in the sensing range start moving in the same direction could lead to a wrong network topology generation. Also, if acceleration correlation is continuous then two trains traveling in opposite directions could have the same acceleration relative to each nodes frame of reference. Secondly, in the ETALON scenario, the devices will be fitted on the train bogies, in this area there are significant vibrations which would introduce a lot of noise in the detection of the comparatively small variation in the bulk acceleration of the vehicle. This means that the detection of the bulk acceleration would rely on filtering of the signal which would introduce a significant computational (and therefore power) overhead and would unlikely be accurate enough for reliable correlation. An additional issue is that there are longitudinal dynamics within a train formation due to tractive and braking forces, track geometry and layout, and the coupling forces. This means that even if acceleration is measured accurately, the acceleration of individual vehicles in a train can vary significantly, particularly in the early stages of movement until the forces reach steady state. Also, there are potentially several situations where the accelerations on two different trains would correlate, producing a false correlation result, this would create false loss of train integrity detections, particularly if correlation was carried out continuously or intermittently after initialisation.

The state transition diagram may be too simplified, a major number of states is probably required to correctly generate the network topology.

The University of Twenty research carries on with the development of a prototype and a test scenario. The prototype has a battery-based power supply, so the device can always be activated independently from the train movement. Since in the ETALON scenario devices need to be powered from energy harvesting devices, it is not always possible to harvest and store the required amount of energy for powering the devices continuously, even when the vehicle is not in motion for significant periods of time. For example, the University of Twenty research starts the network topology generation when the train starts moving, in this time lapse the energy harvesting devices may not have built up enough energy to perform the same actions.

3.3.3 Train Integrity Monitoring Solution

The School of engineering of Florence in 2015 presented a paper that studies a method to overcome the issues imposed by the ETCS L3 [5].

This paper is of interest for its focus on freight train applications, taking in to consideration the limited power supply that is present on this kind of train. The solution is based on a wireless communication between wagons, but it also proposes a cabled communication for internal wagon communication, so the network is still not fully wireless.

The study is based on four wireless devices for each wagon, two are used for communication towards the terminating wagon, while the other two are for communication towards the locomotive. The wireless pair of devices towards the termination wagon are connected through a cabled bus that presents a sensing device in series, while the other couple towards the locomotive are directly connected through a cabled bus.

The sensing device on each wagon is denominated Logic of Vehicle (LdV). This sensing device must be capable to gathering two kinds of information of the wagon: static and dynamic. Static data comprehends vehicle characteristics, while dynamic data comprehends all time-variant variables that want to be monitored (e.g. vehicle state condition).

Regarding the technology employed in the study for this application, three main possibilities have been considered: Zigbee, Bluetooth and Wi-Fi. For the final implementation Zigbee technology was the preferred option, for the higher number of connectable nodes it was able to cope with and for the signal robustness.

Data of each wagon was appended on the message during the communication loop, all this data composes the final message as can be viewed in Figure 7. Some safety measurements have been taken to check the message integrity. Four main information are used for this purpose:

1. A sequence number
2. A timeout packet
3. An identification procedure
4. A safety code

The described architecture was tested by creating an electronic board prototype. Only one board was designed that can cover multiple roles, this differs by the firmware implementation.

At this point a software was designed by the team to test the functionality of the architecture.

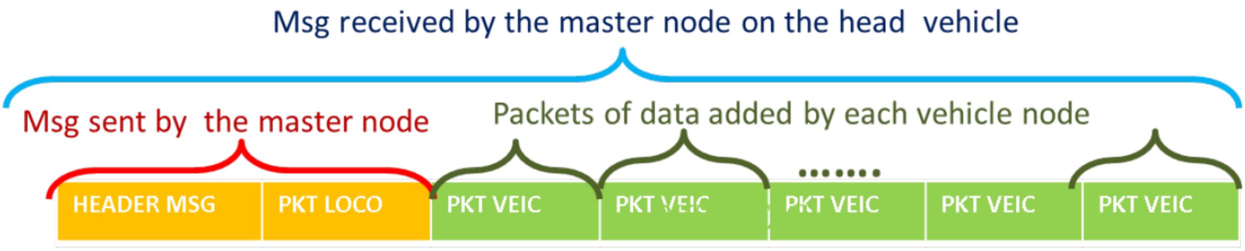

Msg received by the master node on the head vehicle
Msg sent by the master node
Packets of data added by each vehicle node
HEADER MSG
PKT LOCO
PKT VEIC
PKT VEIC
....
PKT VEIC
PKT VEIC
PKT VEIC

Figure 7: Message Structure.

Field	Byte	Description
Nseq	1	Msg sequence number
Len	1	Length

Tot	3	Total length of the message

Field	Byte	Description
Nseq	1	Sequence number
Len	1	Length
LocoID	20	Vehicle ID
CRC32	4	CRC 32
Tot	26	Total length

Field	Byte	Description
Nseq	1	Sequence number
Len	1	Length
VeicID	20	Vehicle ID
UserByte	10	Additional vehicle information (state)
CRC16	2	CRC 16
Tot	34	Total length

Trade-off Analysis for the Train Integrity Monitoring Solution

The research solution mixes a general wireless solution with a locally wired solution, thus still introducing somehow a cabled solution for the communication. While the protocol is explicitly explained, no information is given on how the wireless network topology is generated or how the nodes are aware of each other, probably defining a static network.

Finally, this solution also uses battery powered devices, that also supply the LdV module, that is detached from the RF modules. This approach in not coherent with the ETALON view where each node must integrate the train integrity function, that in this case can be translated in the LdV.

3.3.4 Perpetuum "String" – Linear Mesh Radio Network for Energy Harvester Powered Sensors in Trains

The key properties of a mesh radio network for trains are:

1) The network must be capable of reporting data efficiently when there are thousands of nodes on a single train with a single controller, without the addition of more routers.
2) Network discovery must be rapid, requiring no local intervention.

3) Re-discovery of the network must be automatic when vehicles are moved between locomotives.
4) Unsolicited traffic must be possible – for example, to report a high temperature due to brake or bearing failure.
5) It must be possible for many networks to exist together, without losing synchronisation or data.
6) Network routing must be robust to short term changes in available transmission path.
7) Priority messages must be transmitted more quickly than normal sensor readings.
8) Data packets must be encrypted.

In addition, if the network is harvester powered, the following must also be respected:

1) Average power consumption must, over the course of an operating cycle, be lower than average power harvested.
2) Device reporting times must be variable, to accommodate the power available.
3) Routing must automatically level power consumption across the network.
4) Non-harvester powered sensors may also use the network, but with less overhead (i.e., exploit harvester power to increase lifetime of battery powered sensors-this is optional, but useful).
5) All messages must be acknowledged locally.

To answer these needs, Perpetuum has developed a harvester powered sensor network for trains, using a custom message control layer that could be compatible with a number of radio hardware solutions. In order to meet the need for variable topology, rapid network discovery, adaptation to changes in transmission paths, automatically levelling of power consumption the following major features were developed to deliver a flexible and fault tolerant network solution.

1) Timing synchronisation pulse
    a. Every 10 seconds, a 16byte packet (3ms long) is transmitted from the locomotive mounted, permanently powered data concentrator (network manager).
    b. This timing packet has the data concentrator address, the time, some train data (speed, status) and a hop count.
    c. As each node receives the timing packet, it records the hop count, then (if it doesn’t detect another node transmitting it), transmits the timing packet with an increased hop count.
    d. Each node includes and random hold-off and listen before transmission to resolve contention.
    e. The timing packet is transmitted all the way to the end of the train, each re-transmission is supported by typically 8-12 nodes, depending on transmission/reception efficiency.
    f. By the end of each timing packet cycle, every node has a hop count, that can be used to identify which direction messages are coming from (closer or further away from the network manager).

   g. Typical time for the timing packet to travel down a freight train is expected to be << 1 second (6ms turnaround time for each hop).
   h. All nodes stay powered for the timing packet and synchronise the next power-up time to just before the next packet.

2) Data packet reception/forwarding is handled in a similar way to the timing packet
   a. After the timing packet, the network is open to generate and forward data packets.
   b. Packets can travel towards or away from the network manager simultaneously. The packet direction, address and hop count are the only routing information required.
   c. Packets may be re-transmitted multiple times in poor reception conditions – this does not matter. Redundant transmissions are eliminated as re-transmissions are picked up and the messages removed from the queue.
   d. Data packets are only 64bytes long, with an encrypted data packet and network overhead. Longer datasets can be split between several packets.
   e. Routing is spread randomly between nodes, through the random hold-off mechanism.

Any other sensor can power up, wait for a short time before the timing packet is detected, transmit its data, wait to detect the re-transmission, then go back into low power mode for any arbitrary length of time. The packet will be delivered.

The rate of fresh data generation is relatively low compared to the rate of packet forwarding – this is necessary, because of the topology of the network with many nodes passing data down through a few points. This is made possible by the relatively high-power output available, and the slow rate of deterioration of most rolling stock assets.

All data passes through a multiply redundant route – any node can fail without affecting network reliability. Power consumption by neighbouring nodes will increase slightly.

Nodes only receive/transmit for a short time after the timing pulse is received. Other networks operate at different times, so there is no crosstalk.

The time between timing pulses is short, so the time to re-acquire the network is also short.

Train integrity could be implemented by transmitting a returning timing pulse from the end node (tail light), with additional train connectivity data, before other data transmission commences. This could be adjusted to happen every 5 seconds, since the energy required is very small. Only one of the timing pulses would precede data transmission.

This protocol is characterised by its simplicity, low overhead and low power consumption, although power consumption is not minimised – with sufficient harvester power this is not necessary. Perpetuum believe that the combination of network features, and some novel elements merit protection and have applied for a patent, with the intention of making it an open protocol in the future, permitting many manufacturers to add sensors to the harvester powered data bus. Other undisclosed features will prevent unauthorised use of the network.

WBN could be used to establish the correct consist content before the train moves.

Trade-off Analysis for the Perpetuum "String" – Linear Mesh Radio Network for Energy Harvester Powered Sensors on Trains

The proposed Perpetuum solution is based on proven technology for wirelessly obtaining sensor data from bearings on passenger vehicles, the proposed solution has been adapted to take into account the characteristics of freight trains and the inclusion of train integrity data. This proposal addresses many of the issues identified in the analysis of other systems proposed or considered for train integrity detection, including network topology and power supply. The addition of train integrity functionality through either network analysis or wireless sensors dedicated to detecting a parameter related to train integrity could produce a system suitable for performing train integrity functions.

For the ETALON train integrity case, one important issue and variation would be in relation to establishing the required network topology correctly and to exclude nodes relating to other trains. For a condition monitoring system, the route of data from different trains is not significant, it is not an issue if data from one train is received by another and forwarded to the monitoring centre by that other locomotive. However, for a train integrity function the re-discovery of networks should be managed to ensure that it only takes place when the train is intentionally reformed and not during a journey.

3.3.5 Railway Hazardous Monitoring Solution

The School of Electronics and Information Engineering of the Beijin Jiaotong University presented a paper [6] where WSN are exploited to monitor hazardous objects on the railway. The concept is based on applying a WSN on the train carriages so that railway can be monitored from the dynamic location of trains instead of static monitoring by using sensing nodes directly on the railway.

The part of interest of this work is the implementation methodology for the communication between nodes. The system is battery based and is composed by two main entities: the WSN and the Sink. The WSN communicates to the Sink, transmitting the monitored values, that then sends the data to the DAA Subsystem (Data Acquisition and Analysis Server) together with the position of the train gathered thanks to the GPRS module, the data is sent from the Sink to the DAA using a GSM connection. The architecture overview is depicted in Figure 8.

The WSN communicates through RF modules jumping from node to node to then reach the Sink. The RF communication is implemented using the cc2430 core chip, designed for the IEEE 802.15.4 and ZigBee applications. The routing of the WSN is star-typed based, where the centre of the star is the Sink node, this kind of routing minimises hop between nodes to reach the Sink.

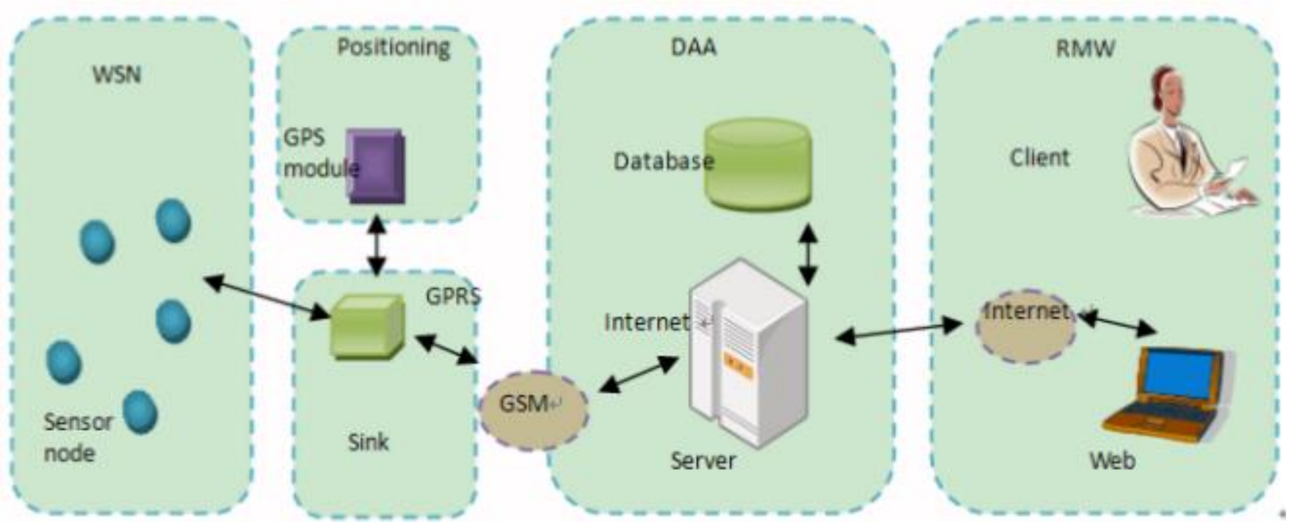


Figure 8: Communication Architecture.

To increase power efficiency technologies of cycled sleep-and-awake and optically coupled isolator controlling mechanisms have been used. Particular attention is focused towards the power consumption of the GPRS module, this is the module that requires more power and also requires the higher voltage supply, so it should be always off during moments where it is not required to read the position of the train.

Trade-off Analysis for the Railway Hazardous Monitoring Solution

The work is of interest to understand how a WSN can be tailored to support communication along a train without having a direct power supply. In the ETALON project a similar network can be applied, where the sink node and the GPRS module are located on the locomotive, where a direct power supply is available, while the nodes can be located on the wagons where no direct power supply is available.

This work though doesn’t take in consideration using energy harvesting technologies to power the WSN, whereas batteries require these being changed after a certain period. Also, the star-typed network is probably not the most applicable routing technology, for the train characteristic the network will be linear and since the Sink node will probably be located on the locomotive, thus meaning that the nodes further away will be unable to reach the Sink without hopping through various nodes.

3.3.6 Reliability Experiments for WSNs in Train Environment Solution

The Uppsala University in Sweden experimented the reliability of WSN in harsh environments, like train environments [7].

The study is focused on two main issue that are present when using WSN in train environments: the wave propagation characteristic around the train and the possibility of energy harvesting to power devices. Firstly, the wave propagation issue is analysed. The experiment is consisting of an antenna in a fixed position transmitting and another moving antenna receiving held by hand; the transmitting antenna is fixed on different parts of the bogie, that correspond to the position of the sensor. The RF communication in established using a 434 MHz frequency using two dipole antennas. Three different polarizations (for the x, y and z axis) have been analysed for the receiving antenna, while the transition antenna is always y-polarized, this path loss can be seen in Figure 9. When moving away from the transmitter the x-polarized measurement has higher fading than the other cases.

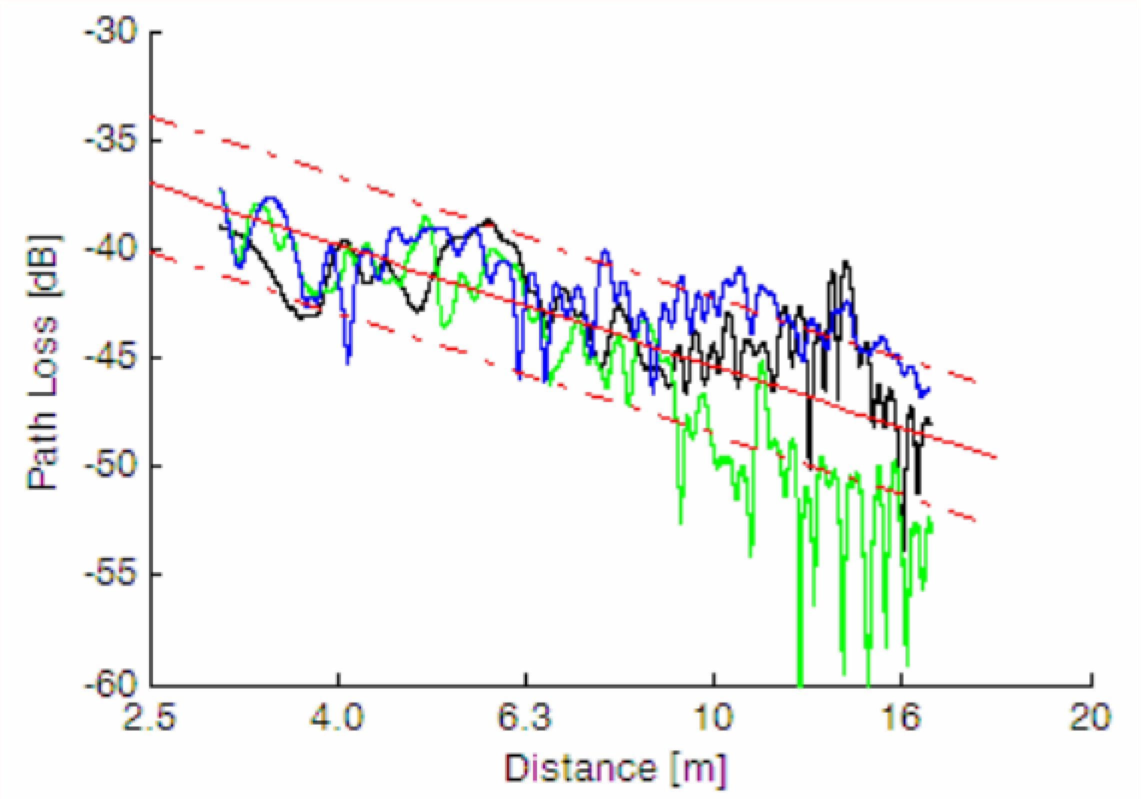


Figure 9: Path Loss for three different polarizations of the antenna - x-polarized in green, z-polarized in black and y-polarized in blue.

Various energy harvesting methods have also been analysed, these include: solar power, vibrations and induction from magnetic fields. The first solution, using solar power, has the main drawback of not harvesting a constant quantity of energy, since this depends on the amount of sun light in the day, also this solution presents the dimension drawback, panels need to have a considerable size. Also, panel need to be clean for maximum efficiency, whereas train environments present a high amount of dirt.

Regarding the exploitation of vibrations for energy harvesting, different techniques are available: electromagnetic by induction, capacitive electrostatic and piezoelectric transduction. The most reliable solution results the usage of piezoelectric transduction, that can also deliver a large amount of energy.

The node will be placed in an environment with very strong magnetic fields, where to obtain a 9V energy source a 200.000 turns coils with an area of 0,01m^2 is requires.

The study continues by investigating the signal reliability as well as the mechanical durability of the device, powering some devices with batteries and one with a solar panel. Since train environments are very harsh due to vibrations, the circuit boards are placed inside a shock absorbing material made of foam rubber, whereas the battery is mechanically separated from the circuit board. The antenna is mounted separately and connected with a flexible cable.

The network is not custom based, but it is developed and marked by TNT-Elektronik AB, that comprehends a getaway node that has the possibility of connecting other devices through USB to read the data.

The study shows that the propagation loss depends on the situations, but generally the path loss is lower on top of the train than beside.

Trade-off Analysis for the Reliability Experiments for Wireless Sensor Networks in Train Environment Solution

This work is of interest for the analysis performed regarding the signal reliability in train environments and the possible energy harvesting technologies that can be applied. The best solution for energy harvesting for the ETALON project seems to be harvesting energy through vibrations, since this is the technology that can guarantee enough energy and does not require high maintenance. Another interesting point that emerges from the study is the signal strength dependent on polarization and position of the antenna. For the ETALON project the antenna will most likely be positioned on the side of the train, specifically on the bogie, thus resulting in an area where there is a higher path loss. Another factor to also keep into account is the antenna polarization, this parameter can highly influence the signal strength.

3.4 Trade-off Analysis Summary

Table 3: Trade-off Analysis Table.

Solution	Wireless	Low Power	Latency	Power supply	Protocol	Network Topology
Passenger Multiple Unit	No	-	-	-	-	-
EU Freight Train	No	-	-	-	-	-
USA Freight Train	No	-	-	-	-	-

Amsted Rail IONX	Yes	Yes	-	Battery	Proprietary	Not specified
Trainspotting	Yes	No (possible addition)	14 ms	Not specified	802.15.4 (xBee)	Linear
Sensing Train Integrity	Yes	Not specified	-	Battery	802.15.4 (xBee)	Linear
Train Integrity Monitoring	Yes - partially	RF module has low-power feature	Few seconds	Battery	802.15.4 (ZigBee)	Linear (EtherCAT)
Perpetuum Energy Harvester Powered Communication	Yes	Yes	-	EH (possibility of battery)	Not specified	Linear
Railway Hazardous Monitoring Solution	Yes	sleep-and-awake and optically coupled isolator	-	Battery	802.15.4	Star-typed
Reliability Experiments for Wireless Sensor Networks in Train Environment Solution	Yes	-	-	Battery and Energy Harvesting	Proprietary (TNT-Elektronik AB)	Not specified

4. TRADE-OFF ANALYSIS FOR TRACK-SIDE COMMUNICATION SYSTEMS

4.1 Introduction

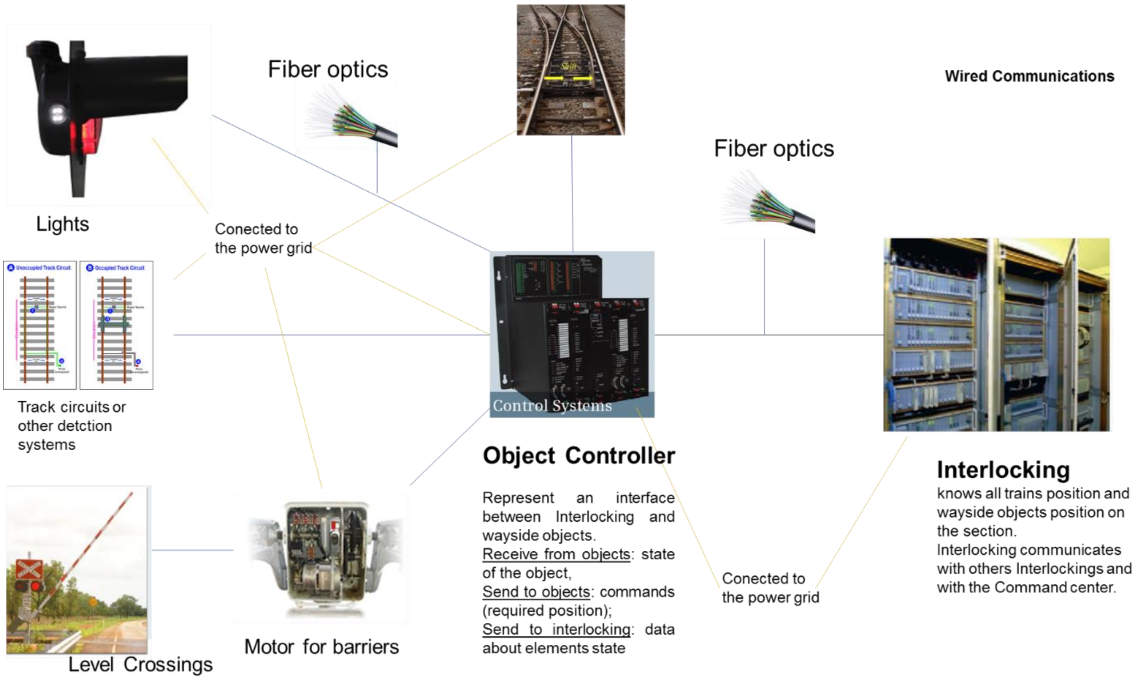
ERTMS introduces, with its level 2 and 3, radio communication intended to substitute current track to train communication based on optical signals and dedicated systems (like track circuits). Level 3 also includes the use of radio communication (integrated by on-board train integrity confirmation equipment) to send to the control centres train location information permitting the elimination of train detection systems. In view of future deployment of ERTMS L3 there is no need of track circuits, signals and axle counters anymore. Balises are used only for providing reference location to the trains.

All communication shall become wireless to decrease installation cost and maintenance, integrating and using real-time status and performance data from the network and from the train, using onboard train integrity solutions and network attached object control functions, supported by wireless network communication

4.2 Current Solutions

4.2.1 Wired Track-side Communication

Currently, all field elements including signals, track circuits, level crossings, switches, variable data Eurobalises are connected using a wired communication to the object controller, from there it goes to the interlocking which processes and forwards the received data to the control centre, and the other way around, the command generated in control centre or by the interlocking are transmitted downlink to the trackside objects.



Object Controller

Represent an interface between Interlocking and wayside objects.
Receive from objects: state of the object,
Send to objects: commands (required position);
Send to interlocking: data about elements state

Interlocking

knows all trains position and wayside objects position on the section.
Interlocking communicates with others Interlockings and with the Command center.

Figure 10: State of the Art of trackside communication systems.

The object controller is directly linked to the electronic control devices (normally, interlockings) being a proprietary solution developed according to the specification of each supplier.

In the Figure 10, two types of communications are depicted:

1 - communications between Interlocking (TMS, ATP, etc.) and object controllers realized by fibre optic and SDH protocol.

It also should be mentioned that the interlockings are connected between each other through fibre optic cables as well.

2 - communications between object controllers and trackside objects using proprietary solutions not standardized generally with copper cables. Every supplier uses its own bus and its own protocol due to security and safety purpose (for example: duplicate transmission channels, security protocol, different protocol simultaneous transmission, etc.).

4.2.2 Wireless Track-side Communication Investigated in ETALON

The purpose of ETALON is to research on radio communications able to substitute the existing cabling, and to identify an energy harvester able to power these communication devices.

The communication between control centre (Interlocking, TMS, ATP, etc) and controlled devices on field shall ensure continuous transmission to provide high availability, detect failures and to supervise electric parameters (to forecast possible needs).

These communications shall allow to enable the concept of self-sufficient Smart Wayside Object Controller (SWOC). The soundness of this concept is analysed and justified in the [13].

The User requirements for SWOC are defined in [14] as following:

- Safely critical applications
  - Data bandwidth > 5kb/s
  - Data throughput > 1 kb/s
  - Latency < 100 / 400 ms, depending on application
- Non-demanding, especially non-safety applications
  - Data bandwidth > 1kb/s
  - Data throughput > 40 b/s
  - Latency < 20s, depending on application

The purpose of a smart object controller is to connect wayside objects such as signals, points etc to the inter-locking systems. Ideally, it would be an autonomous, intelligent, maintenance-free smart equipment able to connect with any signalling wayside object and communicating device in the area, for example by radio or satellite, in order to foster overall reduction both of installation and maintenance costs [13].

No specific requirements so far, but it is clear that there will be both critical (for signalling) and non-critical (maintenance) communication. The communications range is expected to be in order of 100 m to 10 km (e.g. up to ½ distance between 2 stations). A bit-rate of between 10 and 100 kbps is expected. A guideline for expected requirements are given below:

- Safely critical applications
- Data bandwidth > 5kb/s
- Data throughput > 1 kb/s

- Latency < 100 / 400 ms, depending on application
- Non-demanding, especially non-safety applications
- Data bandwidth > 1kb/s
- Data throughput > 40 b/s
- Latency < 20s, depending on application

Most long-range connections of field elements to the central device like Object controller or IXL do not cross distance range of 10 km. The majority of connections for such applications are in the range of 100 to 3000 m. Even lower connection distances around 30 m are used for applications like Level-crossing warning the board to controller or Level-crossing annulment circuit.

4.3 Research Solutions

The Smart Wayside Object Controller (SWOC) concept is a new concept for the Railways which will be investigated and developed within Shift2Rail program.

The SWOC is a piece of equipment that is directly connected to the Wayside Objects, on one side, and to the Route Management Systems (Interlocking, TMS, ATP, etc.), on the other side; and to other SWOCs. The SWOC manages control, maintenance and diagnosis data related to the Wayside Object; and may also supply power to them. Such intelligent objects controllers – knowing and communicating about their status conditions – would not only provide opportunities in terms of cost reduction and asset management improvement but also open new ways of railway network information management and control [13].

ETALON will contribute to this concept analysing radio communication technologies and developing an energy harvester prototype to power the communication devices.

There are various wireless communications technologies that could possible comply with the requirements above stated for safety critical and non-safety critical response applications.

Principally, it could be the IEEE 802.11 - IEEE 802.15.4 families and mobile cellular networks (EDGE, LTE, etc.).

The viability of these technologies for the trackside solution shall be assessed from both technical (including power consumption) and economical point of view what will be done during the ETALON project development.

In the present chapter several research solutions will be analysed covering such features as energy efficiency, reliability, availability, security, fault tolerance and fault recovery.

4.3.1 Linear Sensor Networks: Applications, Issues and Major Research Trends

Classification

The Indian institute of technology and the Galgotia University presented a paper [8] that shows an overview of the Applications, Issues and major Research trends of Linear sensors networks.

This paper has been selected as a reference for the structure of the wireless sensor that could be deployed in the trackside.

Railway is a vital component of modern economy being also the transport sector with most robust and high-demanding Safety and Security requirements that will have a strong impact in the trackside solution. On the other hand, considering the length of railway tracks (and their linear nature, LSN’s are emerging as a solution to railway trackside communications.

The classification of linear sensor network has been summarized in the Figure 11. It is evident from the figure that the linear wireless sensor networks can be classified based on topology and on the hierarchy. In this section, both classifications are discussed.

The topology of Linear Sensor Network can be divided into following three categories based upon the node density: thin, thick and very thick.

- Thin LSN follows a linear structure where all nodes are physically deployed in a straight line.
- The thick LSN follows a structure which is partially linear. Sensing nodes are deployed randomly, and the others are in a straight line.
- The very thick LSN follows a structure in which all nodes are deployed randomly. However, the region is narrow and long.

According to the hierarchy of nodes, LSN’s can be classified into three categories:

- In the one-level network, all nodes have same capacity and functions;
- In the two-level networks, there are two types of devices: sensing nodes and relay nodes;
- In the three level networks, there are three types of devices: sensing nodes, relay nodes and dissemination nodes.

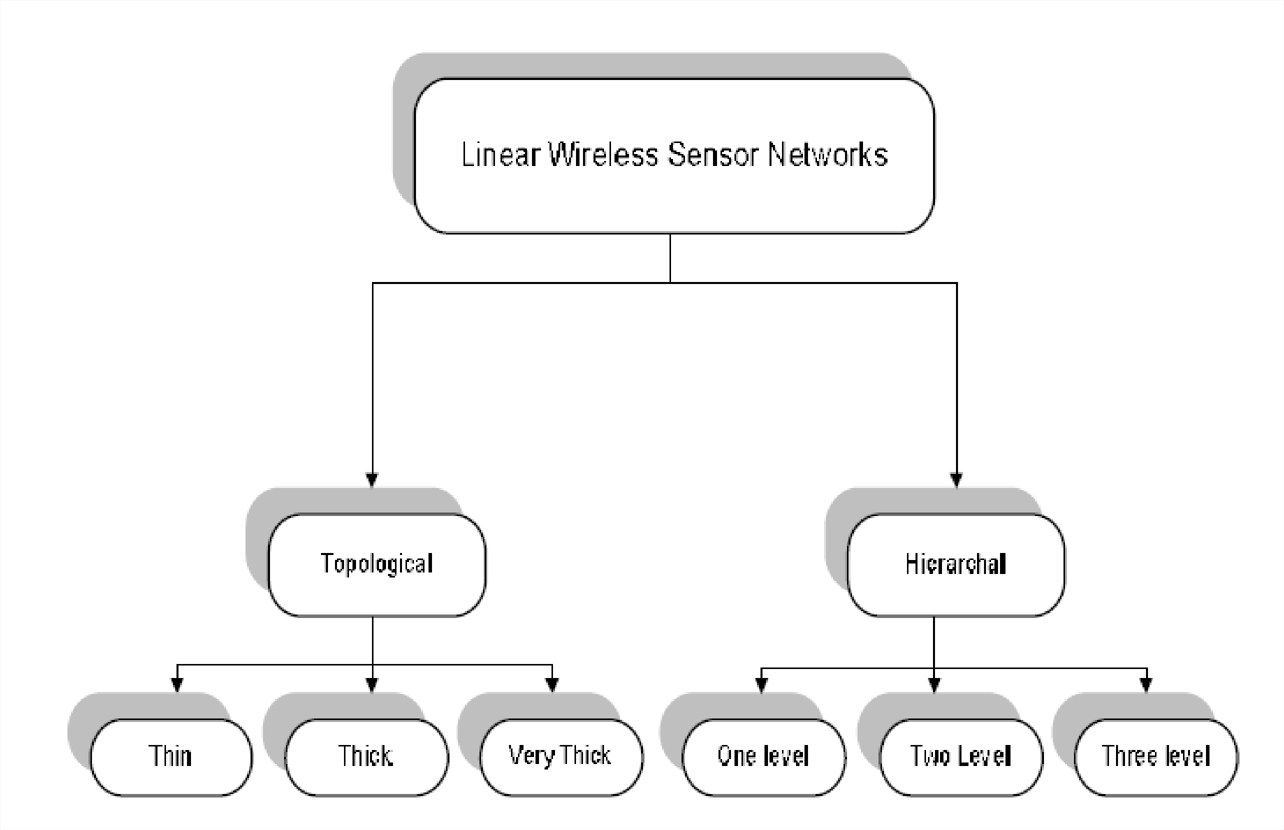


Figure 11: Linear Wireless Sensor Networks.

Sensing nodes are the nodes that are able to “sense” information (as heat, vibration, pressure, wind flow, etc.) from the environment, perform some processing of the sensed information and communicate with other connected nodes in the network. These types of nodes have embedded or connected sensors.

Sensors are classified into three categories: passive, omnidirectional sensors; passive, narrow-beam sensors; and active sensors. Passive sensors sense the data without actually manipulating the environment by active probing. These passive sensors are self-powered, the energy is needed only to amplify their analogical signal. Active sensors actively probe the environment, for example, a sonar or radar sensor, and they require continuous energy from a power source. Narrow-beam sensors have a well-defined notion of direction of measurement, similar to a camera. Omnidirectional sensors have no notion of direction involved in their measurements.

Relay nodes are responsible for routing and forwarding the signal to other nodes, the coordinator or the dissemination nodes

Dissemination nodes: Are nodes able to reach the base station or internet.

Design features

Node Placement: Energy-efficient node placement is one of the main design issues for linear wireless sensor networks. It should be taken in account the spacing between nodes and the range in order to reduce energy consumption per node to a minimum level. The wake up/sleep techniques to achieve a power balanced network should be taken in account.

Topology: it should be proposed an efficient addressing mechanism to allow discovering since the topology used in WSNs is not feasible in LSNs when implementing safety critical applications.

Maximum coverage: In LSNs, area coverage is a problem which impact on the capability of deployed sensors to cover the sensing field. The effective node placement also plays an important role in it. The objective is the optimal placement of networked sensors allowing the total sensing of area under coverage.

Lifetime optimization: If the sensor nodes are deployed in harsh conditions of railway environment with limited energy that may not be easily reloaded/recharged, like in case of remote/rural areas with difficult access, the lifetime optimization becomes a crucial issue. An accepted definition of lifetime of a sensor network is the time span from the instant when the network is deployed to the instant when the network is considered to be non-functional.

Routing Protocols: Routing is one of the most critical issues in any sensor network. Other topics like network lifetime, energy efficiency etc. are heavily dependent on the protocol selected for routing.

Transversal features

Energy Efficiency: In most of the cases, changing the batteries in Wireless Sensor Network is not a trivial task. The alternative to the batteries usage could be the implementation of MAC protocol which provides significant possibilities to effectively regulate the important network performance parameters as throughput, energy efficiency, latency and network lifetime.

Reliability of Network: In non-linear Wireless Sensor Networks (WSN) sinks can be reached through multiple existing paths which enhances the reliability of the network. On the other hand, in LSN's, the number of alternative paths are very much limited. If there are faults in a few contiguous nodes in a LSN, a hole may be created, the network will be partitioned and the nodes in one partition may not be able to reach the nodes in other partition.

Network security: Security plays an important role in several WSN applications.

- Confidentiality: protects secret information from unauthorized entities.
- Integrity: ensures that a message has not been altered by malicious nodes.
- Data Origin Authentication: authenticates the source of message.
- Entity Authentication: authenticates the user/node/base-station is indeed the entity to which it claims to be.
- Access control: restricts access to resources to privileged entities.
- Availability: ensuring desired service may be available whenever required.

Fault Tolerance: In a critical railway environment a high degree of dependability is required. To be considered dependable, LSNs must follow important characteristics such as reliability, availability and maintainability.

Fault Recovery. Once the fault in the system has been detected, the next thing required is to recover from the fault which has been detected.

4.3.2 Linear WSN in M2M Communications: MAC layer Protocol Comparison.

The Department of Electrical and Electronics Engineering of Ege University presented a paper [9] comparing a specific LWSN protocol(LINE-MAC) vs a WSN protocol(S-MAC) for the use in a LNWS.

Protocols comparison

In LWSNs, directional transmission creates significant latency when the end-to-end distance is very long. To solve this problem, hierarchical architectures of LWSN have been proposed where nodes have different roles: basic sensor nodes (BSN), data relay nodes (DRN) and data dissemination nodes (DDN). In LWSNs data packets are relayed hop-by-hop in a chain-to-one pattern. This creates one of the main problems in LWSNs, referred to as the “relay burden problem”. It results in very unequal energy consumption among the nodes and leaves the ones close to the sink with much less energy, exacerbated by the reduced number of neighbours as compared to general WSN. Therefore, the risk of prematurely terminating the network’s operation is greatly increased. Moreover, long sleep times, i.e. short duty cycle solutions are not acceptable because these nodes have extended relaying functions. Another critical issue is that data delivery is more exposed to failure as compared to general WSNs since packet delivery relies on a more limited number of relay nodes. A single node failure can totally disturb the communication process in the network which is a strong weakness of LWSNs. Despite these challenges LWSNs offer some potential benefits. Since the nodes know their neighbours they can schedule packet transmission ahead and regulate their duty cycle accordingly.

In this work a comparative performance evaluation of two protocols representative of the two categories mentioned above is provided: a very generic WSN protocol, S-MAC and a simple but quite efficient LINE-MAC designed for LWSNs. The major interest is in the regards to the energy consumption and end-to-end-delay.

LINE-MAC is a duty cycle based protocol using low power listening (LPL). Nodes follow a dynamically adaptive sleep-wake-up duty cycle; nodes send series of very short preambles and wait for a preACK from downstream neighbours before sending the data itself. Upon awake a node listens for preambles and if none is detected goes to short sleep. If it receives a preamble it immediately sends a short preACK reply and stays awake to receive the number of packets specified in the preamble. When a next hop neighbour overhears a preACK it can deduct how long the data transmission is and sets its timer accordingly to receive the data during the next hop (long sleep). This mechanism of “forward wake-up” provides both distributed synchronization and allows minimizing the hop-to-hop delay. The fact that preambles are very short allows each node to conserve energy while still communicating with its neighbours. The suggested mechanism also has provisions for varying the number of data packets to be sent. It results in gracefully minimizing (or totally avoiding) the “relay burden problem” since every next node will know how many data packets are to be transmitted and will stay awake for the required transmission interval.

LINE-MAC alleviates the three major consequences of the “relay burden problem”: reduces the end-to-end delay, minimizes the consumed energy, and promotes network lifetime by ensuring equal distribution of the consumed energy.

S-MAC, is a well-established MAC protocols for WSNs. It uses a coarse-grained sleep/wakeup cycle, allowing nodes to sleep most of their time. Communication is organized in frames, where each frame begins with a listen period (coordination among nodes that have data to send), followed by a sleep period. In S-MAC, all nodes are free to choose their own listen/ sleep schedules and share them with their neighbours to enable communication between all nodes. Nodes schedule their transmissions during the listen time of their destination nodes. Collision avoidance uses CSMA with an RTS-CTS-DATA-ACK sequence exchange. The low-duty-cycle operation forces the nodes to delay sending a packet until the next listen period of the destination (next hop) which indeed increases latency.

The two protocols were specifically chosen because they have numerous elements in common. Both rely on a sleep-wake up duty cycle to conserve energy. Both adopt a distributed mechanism to synchronize the schedule of neighbouring nodes in order to minimize the transmission delay of a packet travelling from hop-to-hop. However, LINE-MAC benefits from the additional knowledge of who the next one-hop and two-hops neighbours are to ensure longer sleeping times for the nodes and an efficient staggered wake-up pattern.

Scenarios for network topology

The network topology considered is a linear WSN consisting of N, equally spaced nodes indexed by 0,1, 2… (N-1), with fixed communication range. The node with index 0 is always the sink, and all nodes except the sink can sense, transmit and receive, i.e. they are both source and forwarding nodes. All nodes are fixed and know their locations as related to their ID. The density of the nodes is high enough, so that a node can directly communicate with two neighbours on each side. A node connects to the sink directly, either within range or by multi-hop communication. The topology of the sensor network examined in this paper is based on a multi-hop linear network. Each source node generates packets and transmits them through a single-hop or multi-hop communication in order to reach the sink.

The extra level (2-hop transmission) is the simplest hierarchy which allows control of the overall latency.

N1 scenario: all nodes generate packets (with the data they sense) and relay packets from their upward neighbours. Nodes transmit only to their downward 1-hop neighbours.

N1N2 scenario: all nodes generate packets and relay packets, however in this case packets can be sent downwards both to 1-hop and to 2-hop neighbours.

LINE-MAC and S-MAC protocols are implemented in both scenarios to compare the generated energy consumption.

Energy consumption for both scenarios are simulated in Castalia.

In the Figure 15 the consumed energy per packet in relation to the inter-arrival time is presented. Inter-arrival time refers to the time between successive arrivals of the packets of data. It can be deduced that the best performance related to energy consumption of the protocol happens when the successive arrivals of data are around 5s.

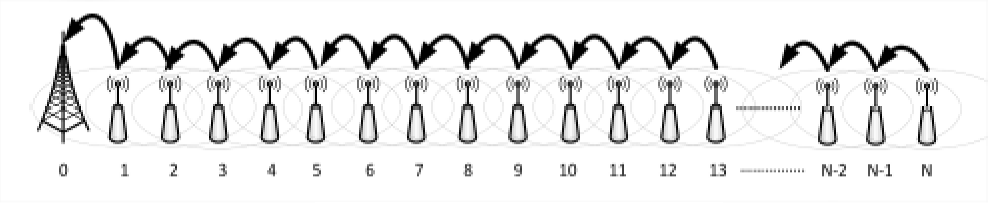


Figure 12: N1 scenario.

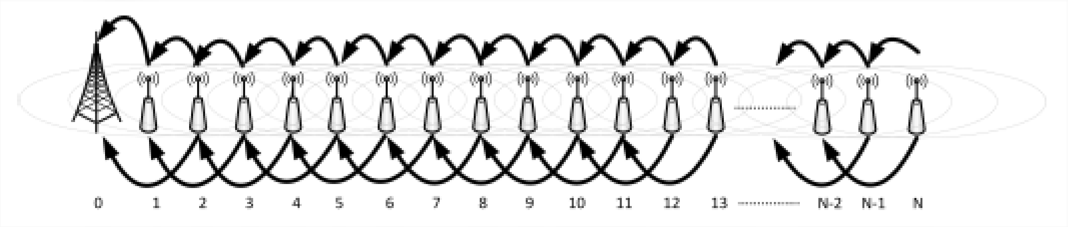


Figure 13: N1N2 scenario.

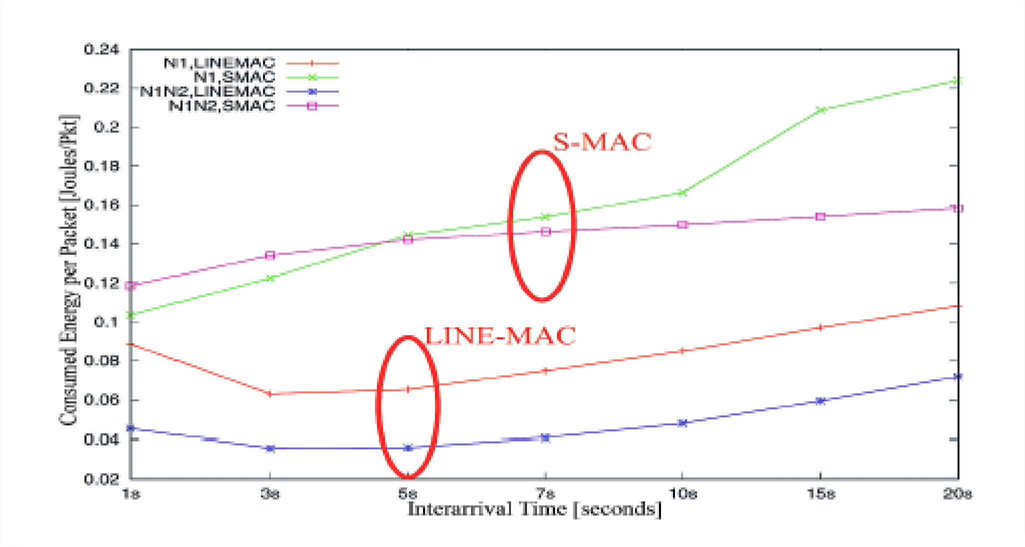


Figure 15: Packet Energy for S-MAC vs LINE-MAC.

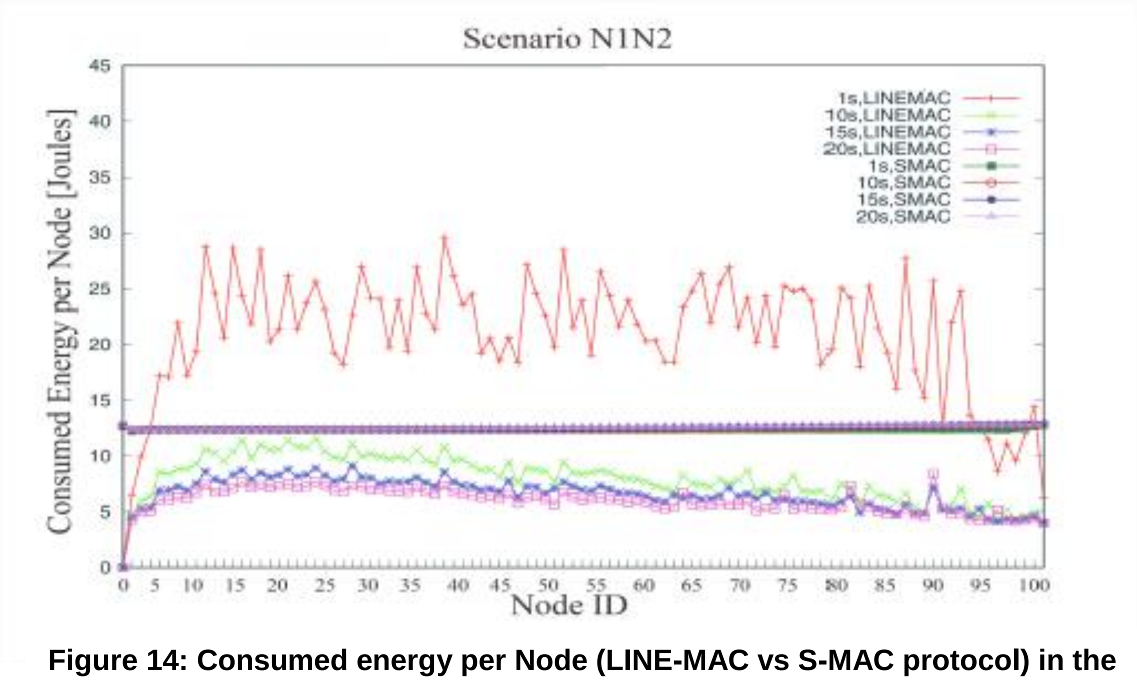


Figure 14: Consumed energy per Node (LINE-MAC vs S-MAC protocol) in the N1N2 scenario.

An important point is that the two protocols have different control overheads (in terms of additional control packets sent), so, in order to provide a fair comparison, the consumed energy is given per packet sent (EC/packet). In both scenarios LINE-MAC is more energy efficient and has less variance in EC/packet. S-MAC exhibits higher energy consumption and range variation for the N1 scenario but is less affected by the traffic intensity in scenario N1N2. Another very important conclusion is that restricting the transmission range (1-hop neighbours) does not provide any energy savings; both S-MAC and LINE-MAC have higher EC/packet for the N1 scenario.

The next figures provide detailed results for the EC/packet per node for the two scenarios. The family of graphs corresponds to different inter-arrival times. The consumption per node for S-MAC does not depend on the scenario, while LINE-MAC shows greater variability especially for the cases of high traffic (1s inter-arrival time).

The performance of LINE-MAC is very dependent on the scenario and N1N2 scenario provides much better results regarding consumed energy per node. An important point is that both scenarios evade the “relay burden problem”, a major reason for depleting the energy of the nodes close to the sink. The “relay burden problem” is a disproportional share of energy consumption between nodes, thus meaning that the risk of prematurely terminating the network’s lifetime is greatly increased.

The Table 4 shows that the packet delivery ratio (PDR) for S-MAC is consistently lower than that of LINE-MAC. Even in the worst case (heaviest traffic) LINEMAC, with only 18.8% of delivered packets, performs nearly 3 times better than S-MAC. However, for light traffic both protocols performance is in the 90% range.

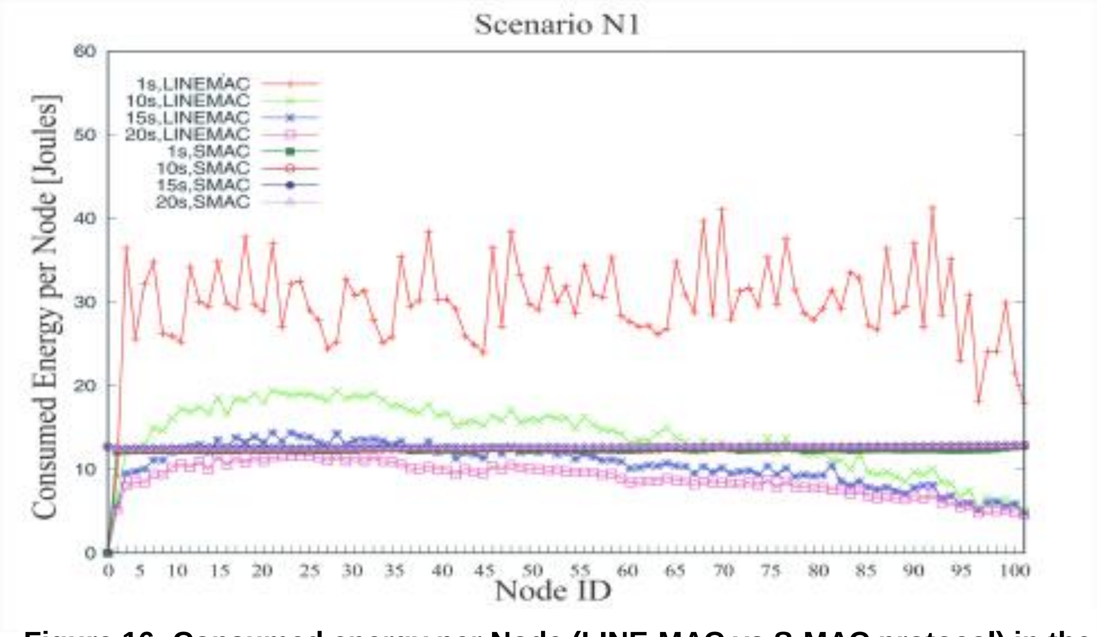


Figure 16: Consumed energy per Node (LINE-MAC vs S-MAC protocol) in the N1 scenario.

Figure 17 shows the Cumulative Density Function (CDF) of latency for both protocols and both scenarios. All latencies are evaluated at the application layer. It clearly shows that LINE-MAC provides consistently lower delays, where most of the packets arrive with up to 5s delay.

Table 4: PDR comparison for scenario N1 and N1N2.

Interval Time	Number of Generated Packets	Scenario N1 Number of Received Packets LINE-MAC	Scenario N1 Number of Received Packets S-MAC	PDR for LINE-MAC	PDR for S-MAC	Scenario N1N2 Number of Received Packets LINE-MAC	Scenario N1N2 Number of Received Packets S-MAC	PDR for LINE-MAC	PDR for S-MAC
1s	179960	33832	12168	18.80%	6.76%	46443	10628	25.81%	5.91%
3s	59960	39002	10293	65.05%	17.17%	43591	9264	72.70%	15.45%
5s	35960	32335	8726	89.92%	24.27%	33834	8788	94.09%	34.44%
7s	25700	23494	8197	91.42%	31.89%	24195	8544	94.14%	33.25%
10s	17960	16502	7578	91.88%	42.19%	16869	8342	93.93%	46.45%
15s	11960	10986	6036	91.86%	50.47%	11235	8099	93.94%	67.72%
20s	8960	8252	5623	92.10%	62.76%	8387	7989	93.60%	89.16%

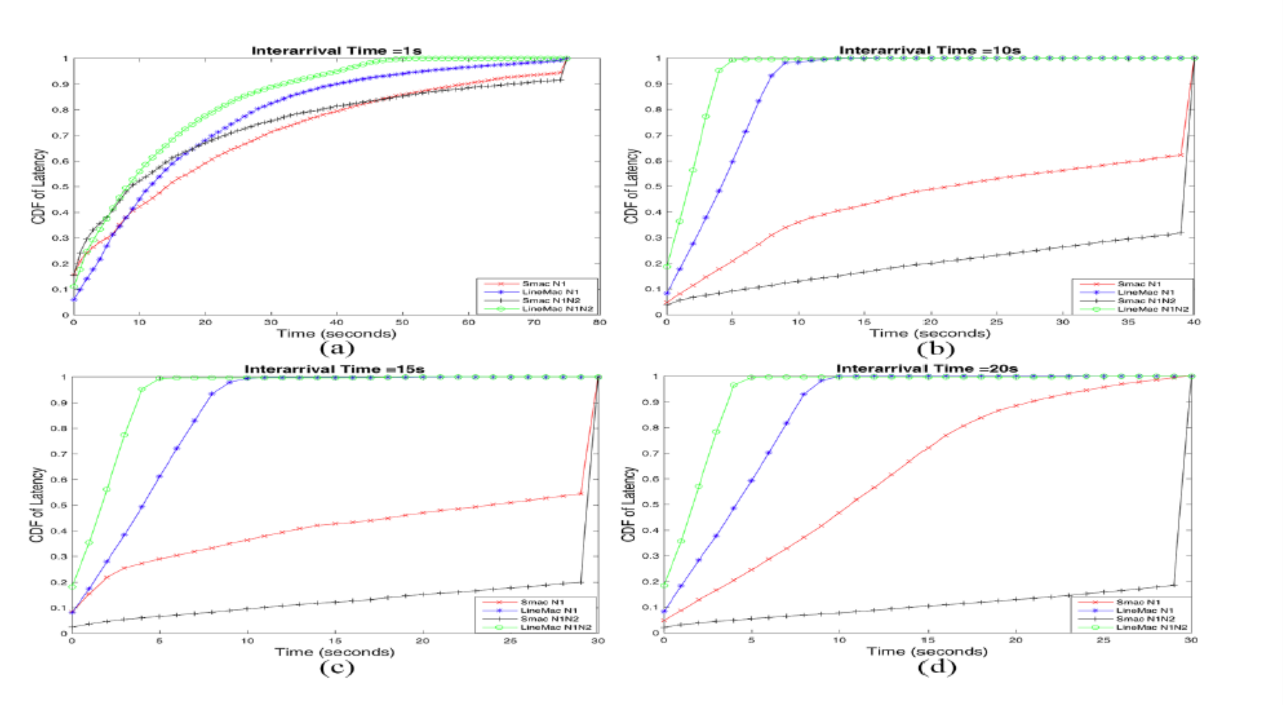


Figure 17: Latency histogram for different interarrival time (a- 1s, b- 10s, c- 15s, d- 20s).

4.3.3 Investigating energy efficiency and timeliness for linear wireless sensor networks

This paper from Ege University, Department of Electrical and Electronics Engineering it compares AREA-MAC and LINE-MAC [10].

LWSNs pose two major challenges: ensuring successful end-to-end delivery and providing a reasonable packet delivery timeframe. The main reason for these is that linear topology limits the number of neighbours and thus the possible transmission routes, so data delivery is more exposed to failure than in traditional WSNs. The authors study the impact of the choice of MAC protocol (TDMA or CDMA) on the behaviour of LWSNs in terms of lifetime and congestion avoidance. The protocols mentioned above improve the performance of the network but require increased node complexity and lack in energy efficiency. The AREA-MAC (Asynchronous, Real-time, Energy-efficient and Adaptive MAC) is a very recent MAC protocol that addresses both time critical and energy-efficient WSN.

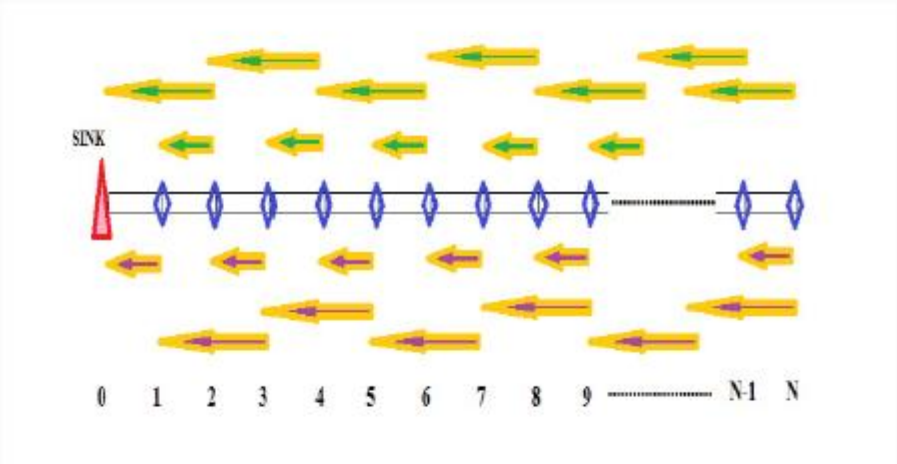


Figure 18: Multi-hop LWSN with N source nodes transmitting to a sink.

The proposed LINE-MAC protocol brings the following three major changes to the operation of AREA-MAC:

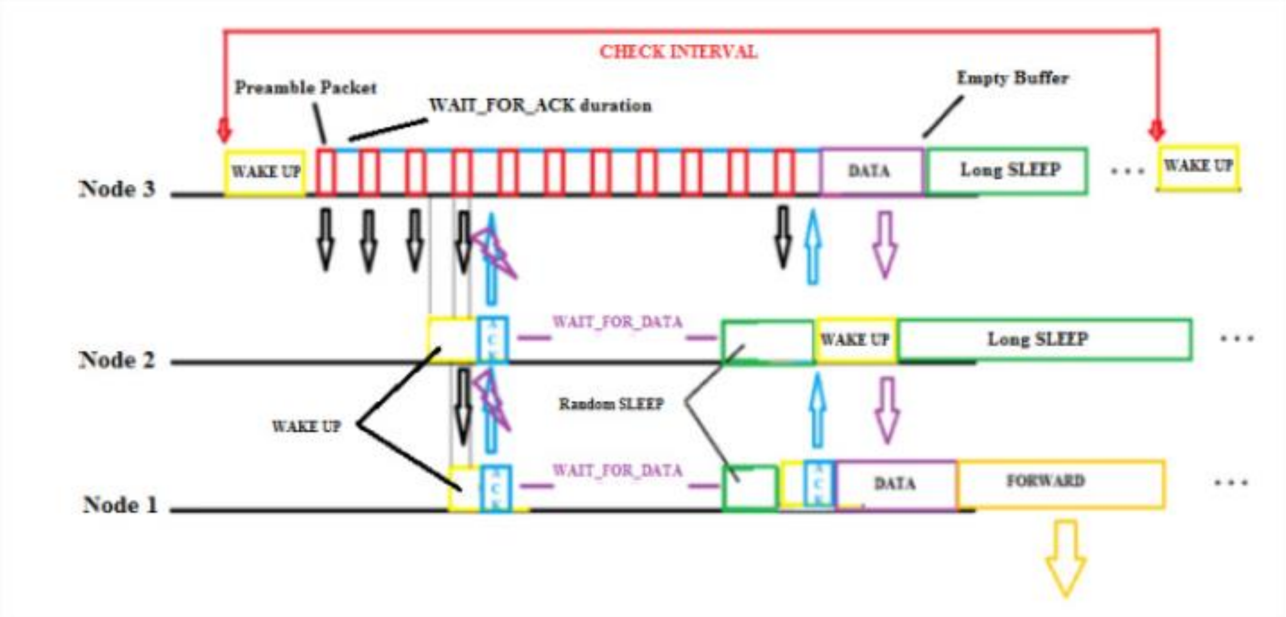


Figure 19: Message exchange in LINE-MAC (part 1).

First: When neighbouring nodes receive broadcasted preambles, they immediately send a preACK packet. Each node activates a timer and if it does not receive any data, instead of entering constant short sleep we propose it enters sleep mode for a random period of time. Thus, preACK collision is avoided because one of the two nodes will wake up earlier and receive the next

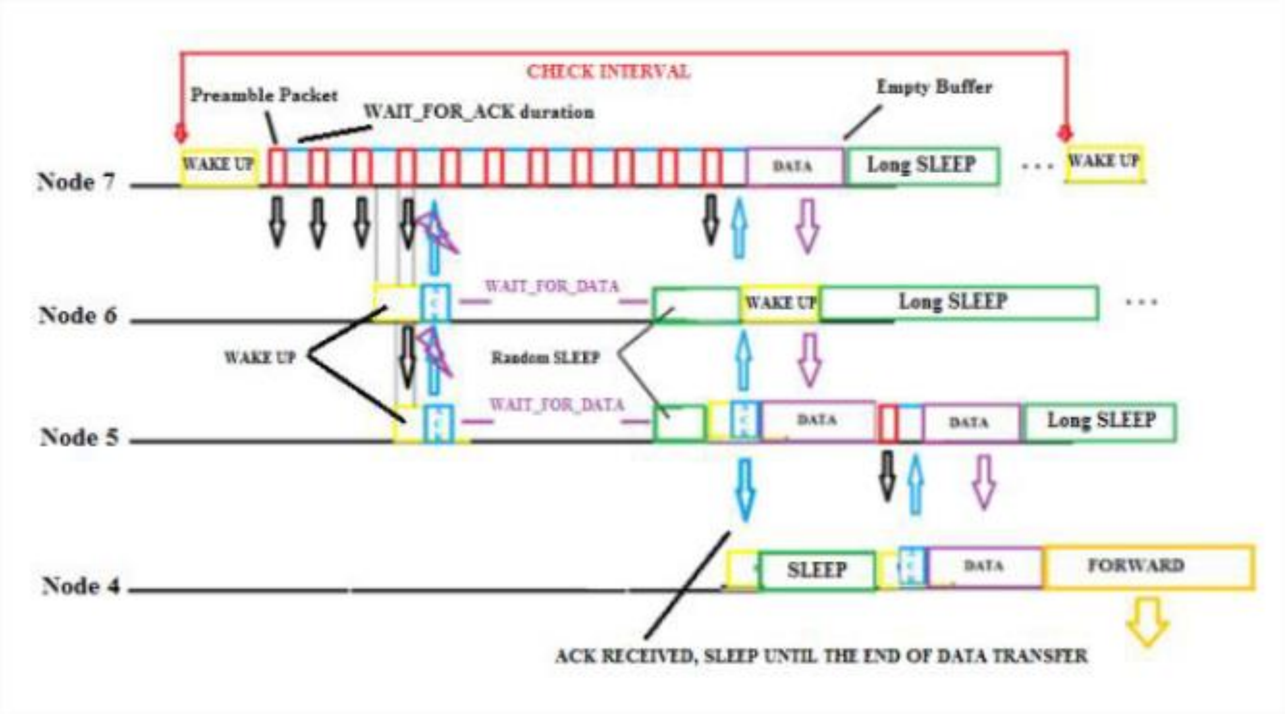


Figure 20: Message exchange in LINE-MAC (part 2).

preamble packet first. The second node, on wake up, finds the channel busy and enters a long sleep.

Second: Nodes closer to sink end up forwarding more packets - “relay burden problem” - which causes higher energy consumption and longer delays. Transmitting only two consecutive data packets is not enough for the nodes closer to sink because the rest of the packets in the queue must wait until the next wake up time. For LWSN this leads to extremely long delays. To reduce the delay, we suggest adjusting the number of consecutive packets to be sent according to the traffic conditions by using the parameter “packetLimit” which can be dynamically changed.

Third: Overheard packets constitute a considerable portion of energy waste in WSNs. For AREA-MAC an overheard packet can be a preamble, an ACK or a data packet. If a node receives a preamble or a preACK, which is not destined to it (overhearing) it immediately enters long sleep mode. However, in line topology, the information contained in the sender/receiver address portion of the preACK packet can be used by overhearing neighbours to determine whether they will be a next hop towards the sink. If an overhearing node is the next forwarding node for the current incoming data packets, it will only go to short sleep until the data transmission ends. When the data transmission of the previous node ends, it will wake up to receive the first preamble packet. Thus, overhearing preACK packets are used to reduce end-to-end delay and mitigate the transmission of unnecessary preambles.

Furthermore, each receiving node sends an acknowledgement to the sender node for each data packet it receives. As we want to adapt the schedule of the overhearing node as close as possible to the on-going neighbouring data transmission process it is important that the overhearing node, which is a prospective forwarding node, knows how many data packets will be sent after the current preACK packet.

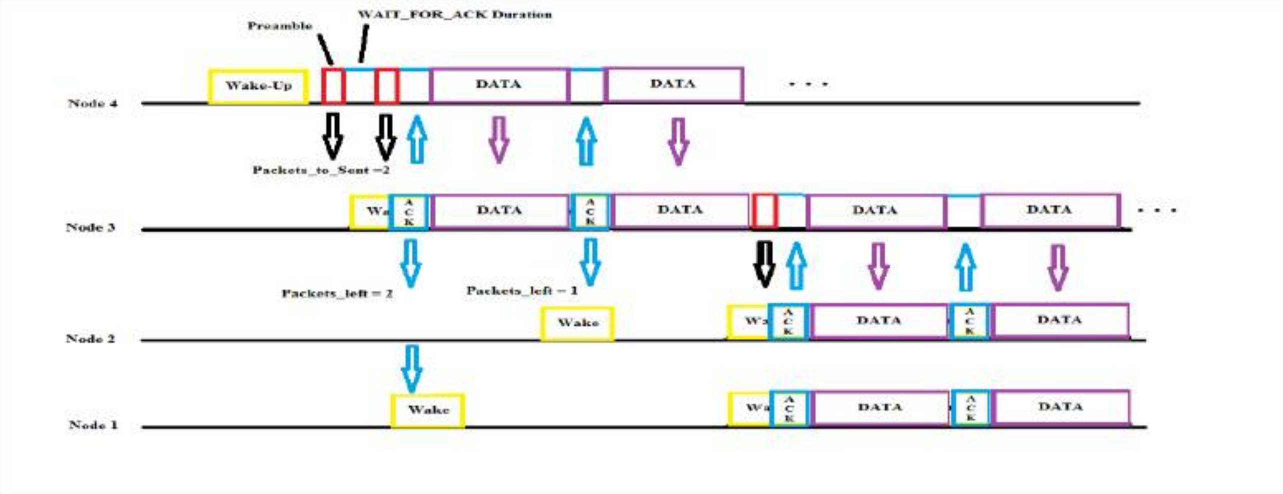


Figure 22: Message exchange in LINE-MAC (part 3).

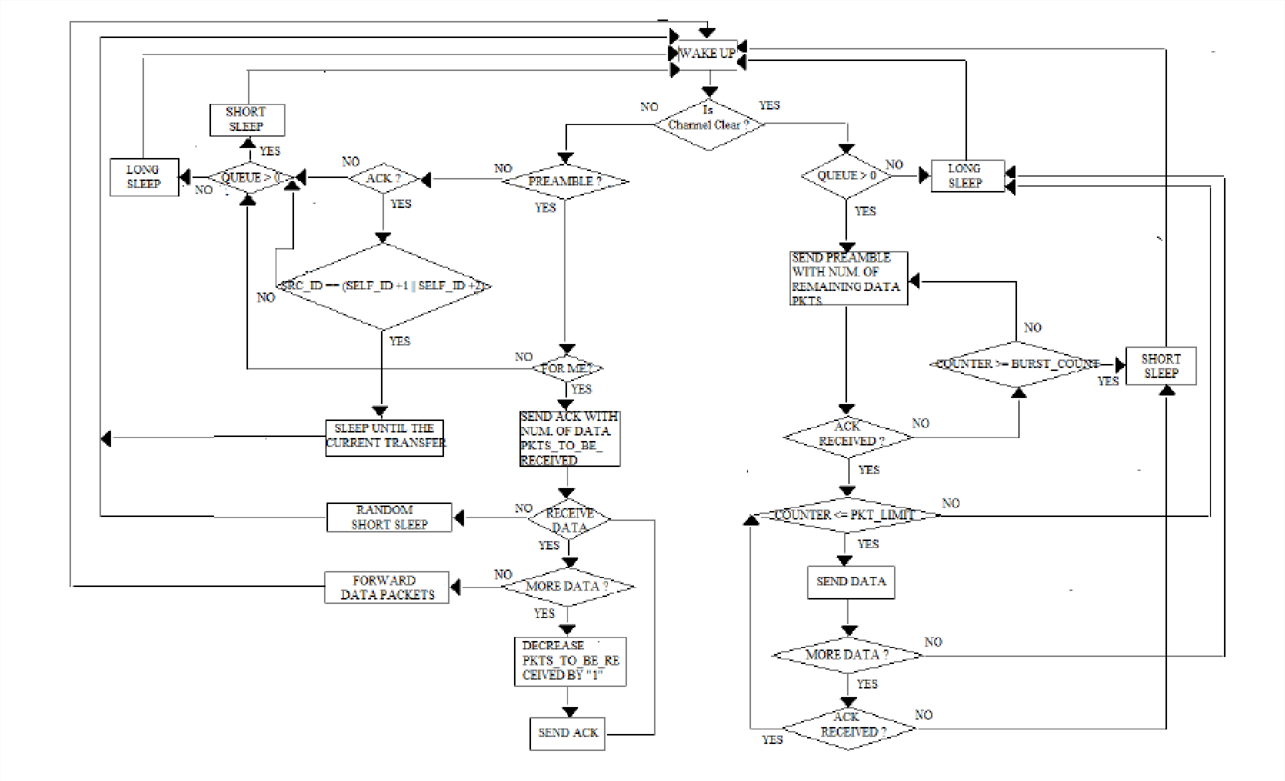


Figure 21: Algorithm of the LINE-MAC protocol.

The algorithm of the whole transmission process using the proposed LINE-MAC protocol is sketched in Figure 22. It improves the timeliness and energy efficiency by using overheard ACK packets and by adding information about the number of consecutive data packets to be sent into the preamble packets.

Second, they evaluate the energy efficiency of the proposed protocol. Figure 23 shows the average energy consumption for each node. For AREA-MAC the energy consumption increases for nodes closer to the sink, which is due to the “relay burden problem. With LINE-MAC, each node consumes nearly the same amount of energy regardless of its proximity to the sink.

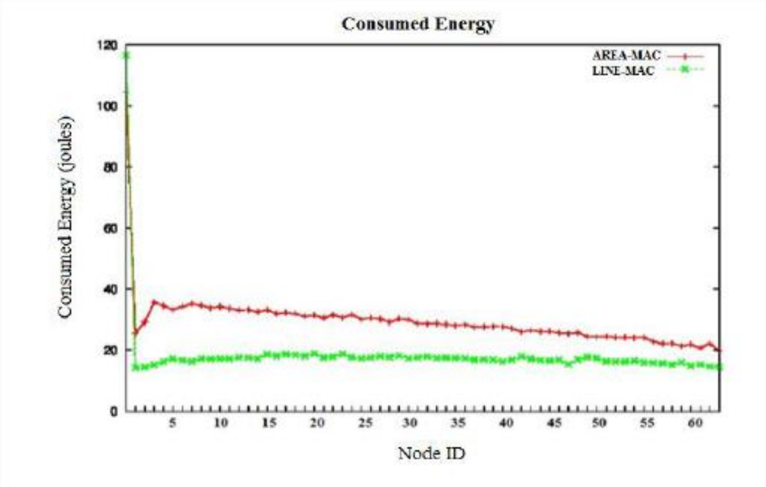


Figure 23: Comparison of energy consumption for linear deployment.

Comments

Relay burden problem: Is a disproportionate share of energy consumption that leaves the “close-in” nodes with considerably less energy. Therefore, the risk of prematurely terminating the network's lifetime is greatly increased

CDMA stands for Code-Division Multiple Access:

- *Code: It refers to the string of binary sequence that the transmitter and the receiver share. This code encodes the information into a low frequency signal before it is transmitted over a channel. This same code is used by the receiver to decode the information. The receiver attains the code with the help of the nearest base station.*
- *Division: In CDMA a single channel is divided into numerous slots which can be used by multiple users. This is possible because of the use of unique code.*
- *Multiple Accesses: Due to code based communication, multiple users can communicate and access the same channel simultaneously without any undesirable interference and loses.*

The frequency of the transmitted signal is made to vary according to a defined pattern (code), so it can be intercepted only by a receiver whose frequency response is programmed with the same code, so it follows exactly along with the transmitter frequency.

Time-division multiple access (TDMA) is a channel access method for shared-medium networks. It allows several users to share the same frequency channel by dividing the signal into different time slots. The users transmit in rapid succession, one after the other, each using its own time slot. This allows multiple stations to share the same transmission medium while using only a part of its channel capacity.

4.3.4 A Distributed Topology Discovery Algorithm for Linear Sensor Networks

This paper is from Faculty of Information Technology UAE University [11]. They show a suitable discovery protocol for LNWS

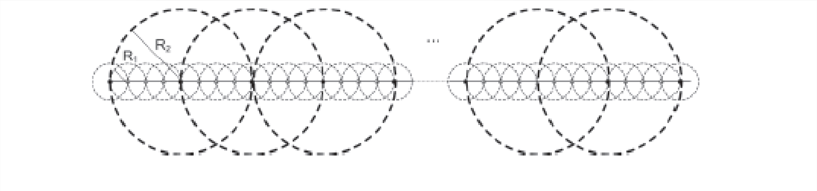


Figure 24: Placement and transmission ranges of the level-1 and level-2 nodes.

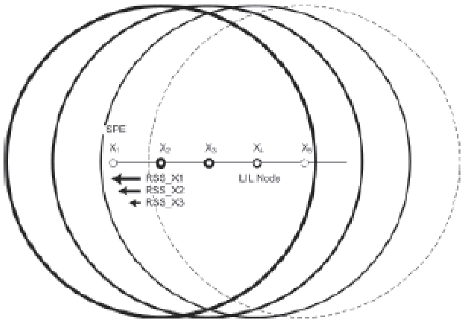


Figure 25: Level 1 topology discovery.

Example of Topology Discovery Process: As illustrated in Figure 26, the SPE (Segment Primary Edge) node broadcasts a topology discovery hello (TD HELLO) message, with an empty L1 list, and a transmission range Rh 1 which is received by x2, x3, and x4. The latter nodes respond by sending TD ACK messages which includes their physical addresses back to the SPE node after a random delay period which is used to avoid collision, in case of collision a retransmission should be performed. The SPE node receives the TD ACK messages, saves their respective received signal strength (RSS), and inserts the physical address of each of the nodes into it in an order which is inversely proportional to the received RSS. Here, it is assumed that RSSx2 > RSSx3 > RSSx4 so the updated list will contain < PHY ADDx2, PHY ADDx3, PHY ADDx4 >. Afterwards, the SPE node unicasts a continue topology discovery hello message, TD HELLO CONT, which contains the updated L1 list to the last node in the list (LIL), which is node x4 in this case. Consequently, the LIL node, x4, repeats the same process in the forward direction by broadcasting another TD HELLO message to its forward neighbours with the updated L1 list. After a similar process, the list will contain < PHY ADDx1, PHY ADDx2, PHY ADDx3, PHY ADDx4, PHY ADDx5, PHY ADDx6 >, and node x4 then unicasts a TD HELLO CONT message to the new LIL node in the L1 list, node x7, which carries on this process forward until the SSE (Segment Secondary Edge) node of that segment is reached and inserted into the L1 list. Then, the SSE node generates a topology discovery completion message (TD L1 COMP), which includes the complete L1 list, and sends it back in a multi-hop fashion through all the nodes in the segment to the SPE node. As the TD L1 COMP message propagates, each of the nodes caches the part of the L1 list that contains k_1^{rt}.1 forward and k_1^{rt}1 backward neighbours, which are involved in the next hop determination during the routing process of data messages through the node. It is not necessary for the nodes to cache the complete L1 list, and the partial caching of the L1 list is done to reduce the size of the routing table in each node.

L2 nodes use a discovery that works in a similar manner.

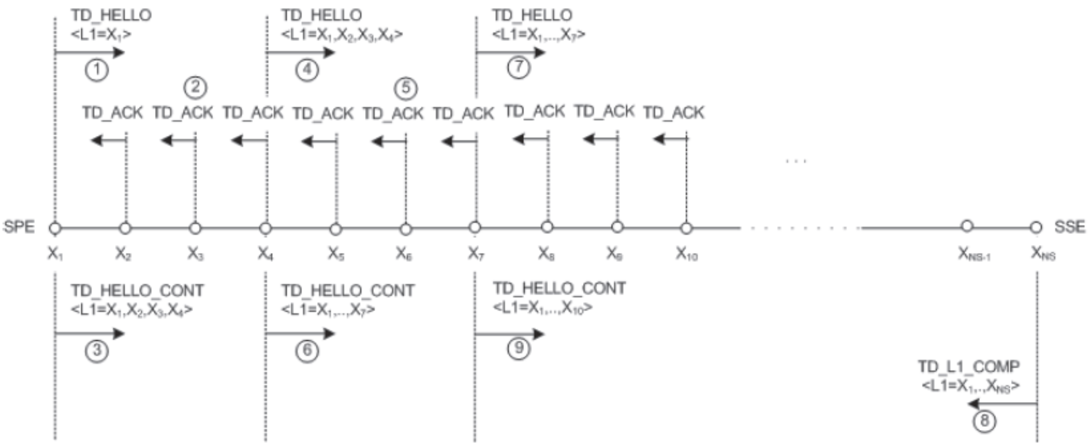


Figure 26: Level 1 topology discovery process.

4.3.5 Modelling the Performance of Faulty Linear Wireless Sensor Networks

This Paper is from the College of Information Technology of UAE University and from the University of Pittsburgh [12].

The distinct way LSNs are designed results in a major challenge of losing connectivity in the network in the presence of multiple node faults [9]. This is due to faults in and consecutive nodes. Nodes can fail due to battery exhaustion, hardware failures, and natural or intentional damage.

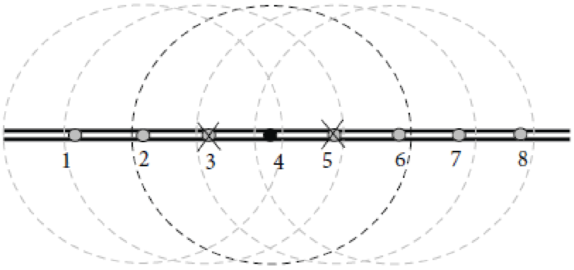


Figure 27: Reliability in a dense LSN.

These consecutive faulty nodes form holes which may cause the LSN to be divided into multiple disconnected segments. Some of these segments will become isolated; thus, they cannot transfer their sensed data to the main station. As a result, the isolated segments will not provide any sensing coverage.

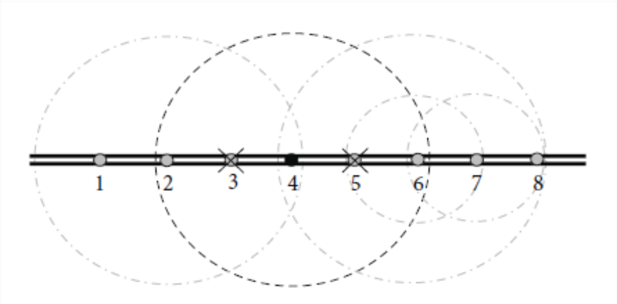


Figure 28: Automatic wireless range configuration.

Each sensor node is usually equipped with a transceiver, a processor, a battery, memory, and small storage in addition to one or more sensing elements. Careful scheduling of these resources is needed to optimize power consumption. Although increasing the transmission range can provide better communication reliability, more energy will be consumed by the nodes. A dynamic configuration for the wireless transmission range can provide better power management. An example of this configuration is in Figure 28. In this network, nodes 3 and 5 have failed. Therefore, the wireless range for node 4 is increased to reach nodes 2 and 6, while other nodes use a smaller transmission range to reduce the power consumption. Connecting both directions to the main station will double the communication reliability.

Fault Type 1-There Are Multiple Holes, Where All Are Not Disconnecting Holes. In this case, the LSN will function normally. Although there will be some holes any hole can be jumped over by extending the wireless range of the node.

Fault Type 2—There Are Multiple Holes, Where Only One of Them Is a Disconnecting Hole. In this case, the LSN will continue to function. This one DH will divide the LSN into two segments: one on the left side of the DH and the other is on the right side. In this case, all healthy nodes on the left side of the DH will use a right-to-left forwarding direction, while all healthy nodes on the right side will use a left-to-right forwarding direction to communicate with the main station.

Fault Type 3—There Are Multiple Holes, Where Only Two of Them Are Disconnecting Holes. In this case, the LSN will be divided into three segments. There are some nodes that are disconnected while the others can communicate with the base station.

Fault Type 4 – There network is isolated, and there are a lot of segments that are completely disconnected.

4.3.6 Securing Wireless Sensor Networks: Security Architectures

There is a critical need for security in a number of applications related to WSN. Resulting from the continuous debate over the most effective means of securing wireless sensor networks, this paper considers a variety of the security architectures employed, and proposed, to date, with this goal in sight.

Authentication is the primary focus, as the most malicious attacks on a network are the work of imposters, such as DOS attacks, packet insertion etc. Authentication can be defined as a security mechanism, whereby, the identity of a node in the network can be identified as a valid node of the network. Subsequently, data authenticity can be achieved; once the integrity of the message sender/receiver has been established

Security protocols strive to be light-weight, in terms of code size and processing requirements, whilst retaining their usefulness, in order to assist the achievability of this goal. To design a completely secure WSN, security must be integrated into every node of the system. Any component of a network implemented without any security could easily become a point of attack. Resultantly, this dictates that security must pervade every aspect of the design of a wireless sensor network application that would collect or disseminate sensitive information.

Conventional networks require protection against eavesdropping, injection or modification of disseminated data packets, and accordingly, most applications of WSNs require the same protection. Cryptography is the standard method of defence against such attacks. This defence brings with it a number of other trade-offs. Varying levels of cryptographic protection implies a proportionately varying level of overhead; in the form of increased packet size, code size, processor usage etc. This is the stem of all debate relating to optimal security techniques in WSNs.

Security in WSNs can be defined as the method of protecting a prospective application against all known types of attack. Attacks including denial-of-service (DOS), traffic analysis, multiple identity/node replication, confidentiality and physical tampering are all areas for concern within WSN security architecture design.

There are many obstacles and constraints involved when designing a security protocol for WSNs. The limited memory, storage and processor capabilities, coupled with stringent power limitations distinguish WSN security architecture design requirements from any other. Harsh environmental operating conditions, the threat of physical compromise and unreliable data transfer also provide challenges for designers.

There are a number of security suites that can be implemented under the IEEE 802.15.4 standard. The most basic can be defined as the secured mode or the unsecured mode (i.e. the null security suite has been chosen). ACL mode provides some limited security services, only allowing the receiving of frames from nodes on the devices Access Control List (ACL). A link layer protocol provides the four basic security services. These include access control, message integrity, message confidentiality and replay protection. An application sets its requirements by setting the appropriate parameters in the radio stack. If there are no parameters entered, then, by default, there is no security enabled. Access control (achieved via the ACL) and message integrity ensures that unauthorized parties should be prohibited from participating in the network. Legitimate nodes should be able to detect messages from unauthorized nodes and reject them. Sequential freshness checks are employed to prevent replayed messages from being accepted by the receiver, as the receiver checks the counter, and rejects any message that has the value equal to or less than the previous obtained counter value. To ensure message authentication and integrity, a Message Authentication Code (MAC) is appended to each message sent. The MAC is viewed as a cryptographically secure checksum of the message. Computing the MAC requires senders and receivers to share a secret cryptographic key, and this key is part of the input to the computation. The sender computes the MAC over the packet and includes it with the packet (using the secret key). A receiver sharing the same key re-computes the MAC and compares it with the MAC in the packet. If the two are the same then the receiver accepts the packet, or otherwise rejects it. Message authentication codes must be difficult to forge without a secret key and, resultantly, if an adversary to the network changes a valid message or introduces a phony message, then it would be unable to compute the corresponding MAC, and authorized receivers will reject any of their attempts to damage the network.

Table 5: Ciphers and authentication methods.

Name	Description
Null	No Security
AES-CTR	Encryption only, CTR Mode

AES-CBC-MAC-128 AES-CBC-MAC-64 AES-CBC-MAC-32	128 bit MAC 64 bit MAC 32 bit MAC
AES-CCM-128 AES-CCM-64 AES-CCM-32	Encryption & 128 bit MAC Encryption & 64 bit MAC Encryption & 32 bit MAC

The standard defines 8 different security suites (TABLE). The security suites can be more broadly classified by their properties. This first of these is the Null suite and provides no security. The next is encryption only (AES-CTR), followed by authentication only (AES-CBCMAC), and finally encryption and authentication (AES-CCM).

Encryption is performed using the AES encryption algorithm, also known as Rijndael. This is defined in the National Institute of Standards and Technology (NIST) Federal Information Processing Standard (FIPS) Publication 197. This algorithm has been a US Government standard since May 2002, and is used by their organizations to protect sensitive information.

AES-CTR (counter mode of cryptographic operation with AES) means that the CTR mode uses AES as the block cipher; and provides access control, data encryption and optional sequential freshness.

Authentication is done using the cipher block chaining with message authentication code (CBC-MAC), which creates a message integrity code using a block cipher in CBC mode, and computes a MAC over the packet and includes the length of the authenticated data. The code can be computed upon packet reception and can be compared with the one received. The IEEE 802.15.4 standard itself includes a detailed description of this process.

AES-CCM is a combination of the encryption and authentication suites detailed above. It has three inputs; the data payload to be encrypted and authenticated, the associated data (header etc.) to be authenticated only, and the nonce to be assigned to the payload and the associated data.

TABLE illustrates that there are varying MAC lengths to choose from for AES-CBC-MAC and AESCCM modes of operation (4, 8 or 16 bytes), allowing for some scalability of security depending on application requirements.

ZigBee is an industrial consortium, which was designed to build a standard data link communication layer for use in ultra-low power wireless communications. The members of this organization came together because they felt that existing standard technologies were not applicable to ultra-low power application scenarios”. The ZigBee network layer (NWK) is designed to operate just above the PHY and MAC layers specified in the IEEE 802.15.4 standard.

The main responsibilities of the ZigBee NWK layer include the mechanisms used to join and leave a network, apply security to frames and to route frames to their intended destinations. The ZigBee specification also details extra security services, including the processes of key exchange and authentication, in addition to those provided under the IEEE 802.15.4, upon which it is built.

Threats of WSN

In order to appreciate the challenge of securing a WSN against attack, it is necessary to consider the possible threats to its security. There are a large and increasing number of threats and attacks to which WSNs are susceptible. They can be broadly classified as attacks against the privacy of the network data or denial of service (DoS).

(DoS) attacks, impersonation or replication attacks and physical attacks. In addition to the types of attack, it is also worth considering that attacks can be launched at any point in the network. This implies that certain attacks may be more effective at different layers of the communications protocol, for example. TABLE depicts the various attacks that can be launched at different layers of the communications stack (similar to the IEEE 802.15.4 / ZigBee; based on the OSI model).

Table 6: Types of Attacks by layer.

Layer	Attack
Physical Layer	DOS – Jamming, Tampering Sybil
Data-link Layer	DOS – Collision, Exhaustion, Unfairness Interrogation Sybil – Data aggregation, Voting
Network Layer	DOS – Neglect & Greed, Homing, Misdirection (Spoofing)(MiTM), Black Holes, Flooding Sybil Wormhole Attack
Transport Layer	DOS – Flooding, De-synchronization

DoS attacks take many forms, as can be seen in the table, and are known to be attacks that can undermine a network's capacity to perform its expected functions. In the case of wireless networks, "jamming" the channel with an interrupting signal is an effective attack, as are flooding or collision attacks, for example. The Sybil attack is composed by a malicious node taking on multiple identities. The node can then launch a number of attacks such as negative reinforcement, or stuffing the ballot box of a voting scheme, for example. This attack is most effective in the higher layers of the communications protocol. The physical security of network nodes is another concern, as it cannot be ensured in certain environments. Nodes are therefore vulnerable to physical harm, or tampering (i.e. reverse engineering). Traffic analysis attacks are forged where the base station is determinable by observation that the majority of packets are being routed to one particular node. If an adversary can compromise the base station, then it can render the network useless.

Node replication attacks can occur if an adversary can copy the node identification of a network node. In this manner packets could be corrupted, misrouted or deleted, and if this adversary could perform this replication it is possible that cryptographic keys could be disclosed. This could be catastrophic for the network. Network traffic is also susceptible to monitoring and eavesdropping. This should be no cause for concern given a robust security protocol, but monitoring could lead to attacks similar to those previously described. It could also lead to 'wormhole' or 'black hole' attacks. These are routing attacks where an adversary convinces a network node of a shorter, or zero, path to the base station, for example, and can disrupt the network in this manner.

Considering the many angles of attack, it is necessary to secure every aspect of the WSN to ensure its successful operation. The requirements of a secure sensor network include data confidentiality, integrity, freshness, availability, autonomy (in terms of organization) and authentication

Security Architectures

The confidentiality and reliability of data, sensed and disseminated, can be vital to the success of many of the applications that WSNs are used for. Data encryption and node authentication are the main defences against attack. There are numerous encryption and authentication protocols available for implementation in WSNs resulting from the continuous development of the Internet. Wireless sensor networks are, more often than not, considered to be one-off deployments of battery-powered, application specific nodes. This dictates that extensive on-chip processing to execute complex encryption/decryption techniques is not a viable option. Optimizing network lifetime is a major goal of research in the area, and accordingly, the most powerful lightweight solutions are sought, as opposed to top-end Internet based solutions that require greater amounts of processing power, in order to realize this objective. In order to achieve a thoroughly secure system, all of the possible attacks must be taken into consideration. These range from eavesdropping on network communication, to denial-of-service attacks and insertion of bogus or malicious packets. Authentication is a mechanism whereby the identity of a node in a network can be identified as a valid member of the network and as such data authenticity can be achieved. This is where the data is appended with a Message Authentication Code (MAC) and can only be viewed by valid nodes capable of decrypting the MAC, through some determinable means. Any messages received from unauthorized network users can be discarded. Traditional broadcast authentication techniques (such as public-key based digital signatures) are not desirable due to the energy constraints on nodes.

The following is a summary to simplify the lecture. For more details about below protocols see the paper [8].

A) SPINS has two secure building blocks, namely Secure Network Encryption Protocol (SNEP) and µTESLA, which run on top of the TinyOS operating system. SNEP is used to provide confidentiality through encryption and authentication; whilst also providing integrity and freshness. µTESLA is used to provide authentication for broadcasted data.An Important evaluation considerations in this type of system include code size and speed of operation. SPINS can occupy from 1580 to 2674 bytes of memory, and 6.30 to 7.24 ms to complete, depending on optimization. The memory size (RAM) required by the modules is 220 bytes. Approximately 20% of the energy cost of sending a packet secured under SPINS is consumed by the security operations.

B) TINYSEC is designed as a replacement for the unfinished SNEP a “Link Layer Security Architecture for Wireless Sensor Networks”. It provides similar services, including access control, message integrity and confidentiality. Access control and integrity are ensured through authentication, and confidentiality through encryption. TinySec allows for two specific variants. The first of these, TinySec-Auth, provides for authentication only, and the second, TinySec-AE, provides both authentication and encryption. For TinySec-Auth, the entire packet is authenticated using a MAC, but the payload data is not encrypted; whilst using authenticated encryption, TinySec encrypts the payload and then authenticates the packet with a MAC. TinySec is implemented in approximately 3000 lines of nesC code. It requires 728 bytes of RAM and 7146 bytes of program space. Completion of the cipher operation (Skipjack) takes approximately 0.38 ms. Depending on the mode of TinySec, authentication only or encrypted authentication, an increase in energy consumption of 3.03% or 9.1% is present, respectively, over that, sending a normal TinyOS packet, can be seen in clear. The increased energy cost can be mainly attributed to the increased packet size induced through the provision of the security functions.

C) LEAP/LEAP+: Localized Encryption and Authentication Protocol management protocol for sensor networks, motivated by the observation that different types of messages propagated in wireless sensor networks have different security requirements. Lightweight, energy efficient operation and robustness and survivability in the face of node compromise, are the main design goals of this protocol. This security architecture now has many similar attributes to its predecessors. It is implemented, written in nesC code for TinyOS, in 17.8 Kb of memory (ROM), and the RAM usage is determined by the number of neighbours present for each node. For one node it uses 600 bytes, for thirty nodes it uses 1.59 Kb. This is reasonable usage based on the 128 Kb of program memory and 4 Kb of RAM provided by the Mica2 mote. In this prototype implementation, a node can complete determination of pairwise keys with three neighbours in approximately 8.5 seconds.

D) Security Manager: It is based on a Public Key Infrastructure (PKI) and Elliptic Curve Cryptography (ECC). The Security Manager (SM) gives static domain parameters such as the base point and elliptic curve coefficients to prospective network nodes. Devices use these initial parameters to establish permanent public keys and ephemeral public keys, which are in turn used for securing the network data. After calculating a public key, a node sends this to the SM, which could have a public key list for all nodes in the network. ECC is an approach to public-key cryptography which is based on the algebraic structure of elliptic curves, over finite fields. Elliptic Curve algorithms provide reasonable computational loads and smaller key sizes for equivalent security to other techniques. Smaller keys sizes reduce the size of message buffers and implementation cost of protocols. Authenticated key agreement is achieved via the SM, based on the EC-MQV algorithm. This algorithm is more advanced than Diffie-Hellman, eliminating the man-in-the-middle attack. Diffie-Hellman is included in the EC-MQV algorithm as a subset. The scalable nature of this solution is attractive for use with wireless sensor networking, as it can be increased or reduced depending on the nature of the disseminated message. A concern would be the number of keys the SM would be required to store as the network grows, and every node may not come into range of the SM. This can be addressed by allowing more than one node to have SM responsibilities. The trade-off between power conservation will still apply, but may be significantly less than other public key architectures.

E) ZigBee: the ZigBee specification outlines the design of the NWK layer that operates just above the PHY and MAC layers specified by the IEEE802.15.4 standard. The concept of a “Trust Centre” is introduced in the specification. Generally, the ZigBee coordinator performs this duty. The coordinator allows other devices to join the network and distributes the appropriate keying Information. There are three roles played by the “Trust Centre”:

1. The trust manager, whereby authentication of devices requesting to join the network is carried out;
2. The network manager, maintaining and distributing network keys;
3. The configuration manager, enabling end-to-end security between devices.

There are two modes of operation; Residential Mode and Commercial Mode. Running the former, low security residential applications are accounted for. The latter is designed for high-security commercial applications. In Residential Mode, the Trust Centre will allow devices to join the network but does not establish keys with the network devices. It therefore cannot periodically update keys and allows for the memory cost to be minimal, as it cannot scale with the size of the network. In Commercial Mode, it establishes and maintains keys and freshness counters with every device in the network, allowing centralized control and update of keys. This results in a memory cost that could scale with the size of the network. There are three types of keys specified for use in ZigBee security services; the Master Key, the Link Key and the Network Key. Master keys are installed first, either in the factory or out of band. They are sent from the Trust Centre and are the basis for long-term security between two devices. The Link Key is a basis of security between two devices and the Network Keys are the basis of security across the entire network. Link and Network Keys, which are installed either in the factory or out of band, employ symmetrical key-key exchange (SKKE) handshake between devices. The key is transported from the Trust Centre for both types of keys. This operation occurs only in Commercial Mode, as Residential Mode does not allow for authentication. The scalability of the security suites in this architecture is ideal for developing the technology, as implementations that do/do not require high levels of protection can be developed using the same platform. It is also useful for applications which may need to send some messages with higher security than others. Implicitly, there will be increases in both the message latency and power consumption of the motes the higher the security requirement. Considering that security is integrated into the ZigBee protocol, it is not worth considering extra code space required for implementation etc. Power consumption and latency are currently under investigation, but these trade-offs in relation to security provided will not prove detrimental to the application and usefulness of ZigBee in the wireless sensor networking domain.

4.3.7 Utilizing Unused Network Capacity for Battery Lifetime Extension of LTE Devices

An extensive research has been performed in the fields of battery design and system optimization with regard on energy consumption. Due to the fact that the radio part of devices is one of the most important energy consumers, a special focus has to be set on different approaches and strategies on how to reduce the energy that has to be spent for the transmission of a certain amount of data.

This paper presents an approach that is based on the reduction of the necessary uplink transmission power while simultaneously adapting the Modulation and Coding Scheme (MCS) as well as the number of allocated Resource Blocks (RB).

Detailed measurement results show that especially those mobile stations which are submitting at a close to maximum transmit power can significantly reduce their averagely consumed power if they can avoid operating the Power Amplifier (PA) in its high-power mode. In many cases this can be achieved by a minor reduction of the transmit power. The decreased signal quality, in terms of a smaller Signal to Noise Ratio (SNR) at the Base Station (BS), must simultaneously be compensated by choosing a more robust MCS and additional RBs. Actually, one can say that the energy savings are paid by radio resources which is acceptable if the cell is not completely occupied.

The transmission power that is used for LTE uplink signals has a major influence on the energy consumption of the radio part of a mobile device. Therefore, the aim of any system designer should be to reduce this figure as far as possible without decreasing the signal quality.

A promising approach for the reduction of the consumed energy is to avoid the transition of the power amplifier to the high-power mode. In most cases this target can be achieved by a reduction of the transmit power which comes along with a decreased SNR at the BS. To avoid signal degradation the MCS has to be

adjusted to allow for a higher robustness of the signal. As the order of the MCS is reduced, the number of data bits that can be submitted at one RB is decreased correspondingly. Therefore, if the data rate shall remain constant the number of allocated RB has to be increased. A specific so-called Transport Bock Size (TBS) is defined by an MCS together with the number of allocated RB. Therefore, the same TBS can be achieved by different combinations of MCS and RB allocations.

Figure 29 illustrates the generic approach for the reduction of the actual power consumption by means of a flow chart. From this one can see that the procedure can be divided into four steps:

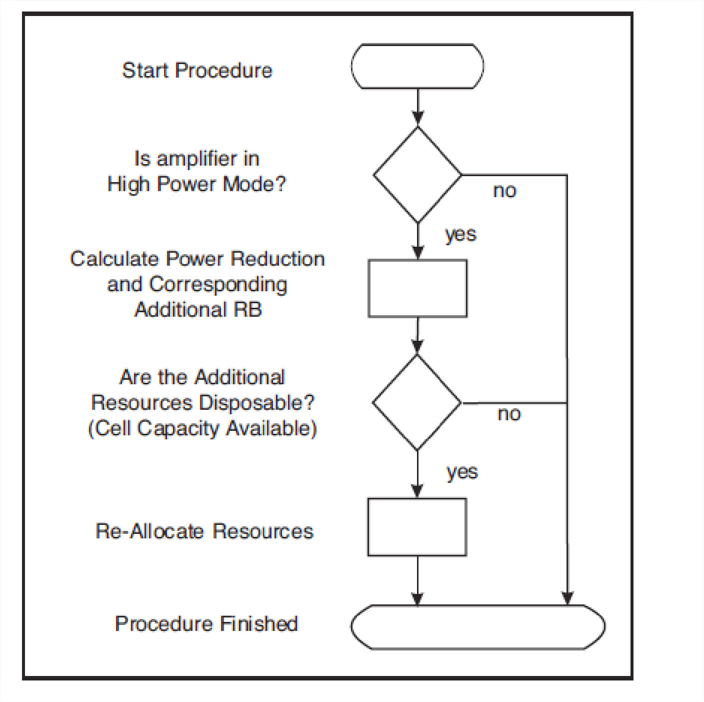

Start Procedure
Is amplifier in High Power Mode?
no
yes
Calculate Power Reduction and Corresponding Additional RB
Are the Additional Resources Disposable? (Cell Capacity Available)
no
yes
Re-Allocate Resources
Procedure Finished

Figure 29: Flow Chart Illustrating the Generic Power Reduction Scheme.

1. Check if the device is currently operated in high power mode, e.g. the transmit power is close to the maximum.
2. If this is true, calculate the needed power savings. From this derive the MCS that is needed to work properly with the decreased SNR. Finally, calculate the number of additional RB needed to compensate for the reduced number of data bits per RB.
3. Check if the additional RB are available in the cell, e.g. the cell is not completely occupied, or the user is considered as priority user.
4. If the additional RB are available, re-allocate the resources and reduce the tx-power for saving accumulator lifetime.

While the information that the LTE device is operated in the high-power mode is only available at the device itself, the resource allocation in LTE does usually take place in the eNodeB. Therefore, a protocol modification will be needed which allows the UE for asking for additional resources for power saving purposes. This could be done by slightly modifying the power headroom reports.

That the final decision about the power saving mode of a user equipment (UE) takes place at the BS comes along with the advantage that the additional RB can be retrieved immediately if new users enter the cell and there is no more capacity available for energy saving purposes.

Figure 30 shows the results of ray-tracing simulations for an exemplary urban LTE cell. The simulation calculates the SNR that can be achieved at the BS if the UE submits at the max. tx power of 23 dBm. One can see from Figure 30 that the users can be divided into three groups: The first group is not able to establish an LTE connection with a BLER below 0.01 because the achievable SNR is below 2 dB. The second group can establish a connection but has to transmit in the high-power mode because a power reduction of 14 dB, as needed for a fall back to low power mode, would cause a connection loss. The third group finally, that is the biggest one with above 50 %, can fall back to the low power mode and therefore achieve significant power savings if additional resources are available in the cell. If the available resources are not sufficient for applying the algorithm to all the candidate UE, it is up to the operator to define priority rules. It is worth noting that the performance of this approach is yet only validated for applications with constant data-rate, e.g. real-time applications such as streaming or interactive applications. If a file transfer has to be performed it might be suitable to minimize the transmission time no matter what the average power consumption is.

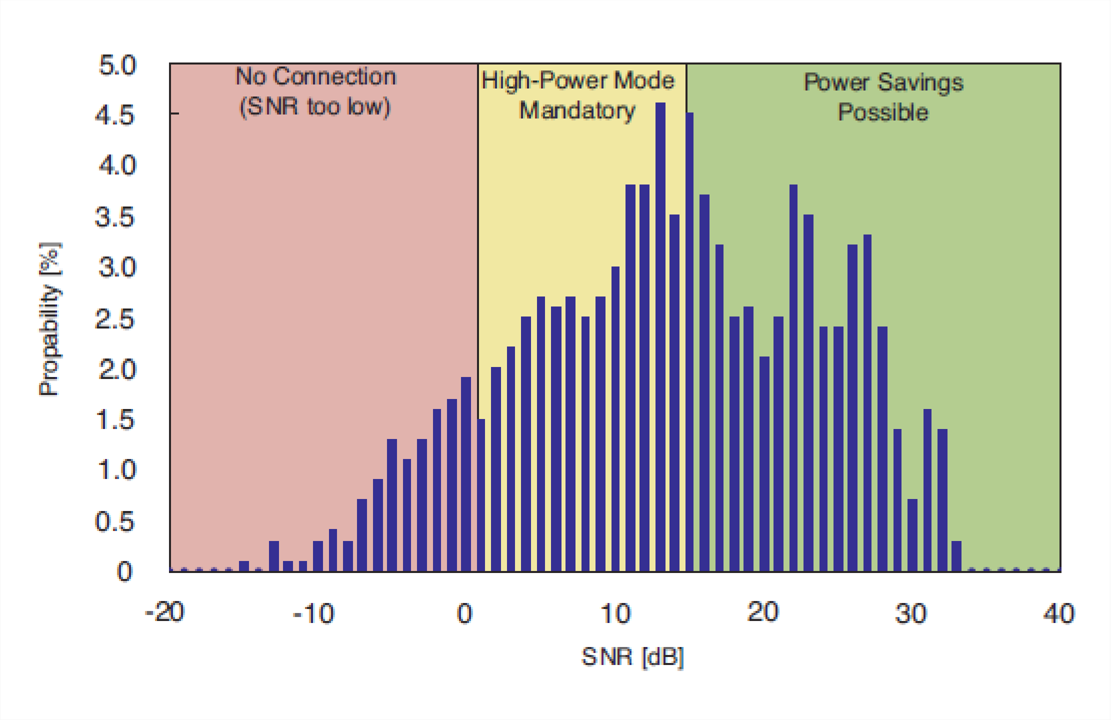


Figure 30: Probability Density Function (PDF) of the SNR for Typical Urban Environment (@24dBm tx-power).

For a concrete quantification of the energy savings that can be achieved by the novel approach a detailed measurement for Voice over IP (VoIP) traffic as an exemplary and prominent real time application has been performed.

The relationship between the SNR and the corresponding block to error ratio (BLER) were derived by means of simulations using OPNET Wireless Suite and the specialized LTE model herein. The resulting plot can be seen in Figure 31. An important figure that can be derived from the plot is the amount of signal power that can be saved if the highest order MCS, which is a 16-QAM (ID 17) with a code rate R = 0.569 is replaced by a more robust MCS. If for example the more robust MCS 11 is used instead of MCS 17 the transmission power that has to be spent for achieving the target BLER of 1 % can be reduced by 5.8 dB.

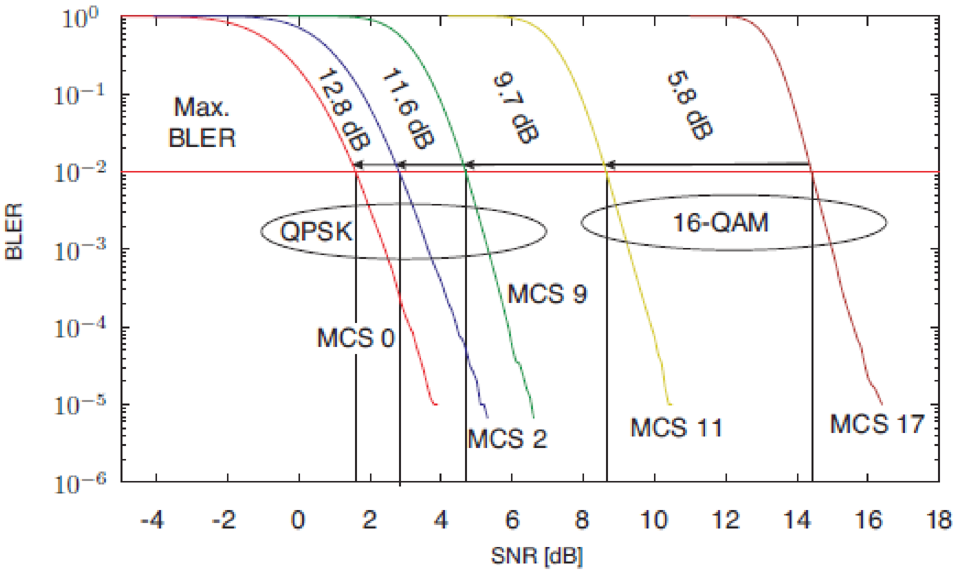


Figure 31: SNR vs. BLER for different Modulation and Coding Schemes.

For the final performance evaluation of the novel scheme the actual average power consumption was measured for the different resource allocations shown in the Table 7.

Table 7: Average power consumption for different Modulation and Coding Schemes.

RB	MCS	R	SNR for BLER < 1%(db)
1	16-QAM (ID 17)	0.569	14.4
2	16-QAM (ID 11)	0.285	8.6
4	QPSK (ID 5)	0.285	4.7
8	QPSK (ID 2)	0.142	2.8
12	QPSK (ID 0)	0.095	1.6

Besides the parameterization of the MCS, the number of RB and uplink transmission power per RB P0, the signal was parametrized that only the necessary fraction of TTI is occupied by a VoIP transport block. Figure 32 shows the average results of the measurement together with the 95 % confidence intervals. As it can be seen from the plot, one second of VoIP communication is costing 2.14 J from the accumulator of the LTE device if one RB with MCS 17 coded data is used for the submission of the data. For that, we assume that for using MCS 17 the transmission power of 7 dBm/RB is needed for ensuring the required BLER. If the same user would switch to MCS 11 it requires one additional RB to ensuring a TBS of 328. On the other hand, the user would be able to decrease the uplink power by about 6 dBm/RB which makes the devices amplifier fall back from high power mode to low power mode. This transition leads to a decreased power consumption of the device to only 1.62 J/s which makes a power saving of 23.8 %.

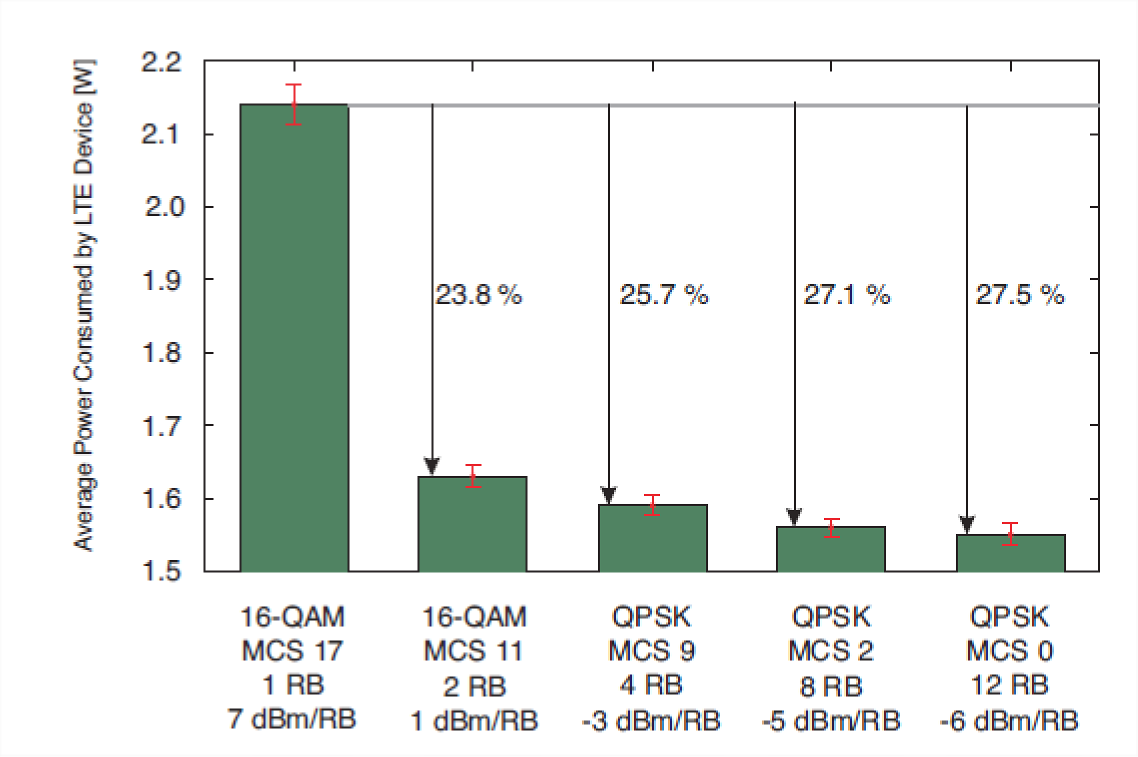


Figure 32: Power Consumption of LTE Device for Different MCS/RB Constellations supporting TBS 328 (VoIP).

4.3.8 An Accurate Measurement-Based Power Consumption Model for LTE Uplink Transmission

The communication activities -in particular the uplink transmissions- are one of the major drivers of the overall power consumption. The important relationship between the uplink transmission power PTx [dBm] and the overall power consumption of the UE, P [W], has however not been sufficiently addressed for LTE so far.

Especially for high transmission powers, the LTE UE platform consumes a disproportional amount of power leading to increased heat loss.

To derive an accurate power consumption model, extensive measurements for different commercially available LTE UEs have been performed. The measured correlation between the uplink transmission power and the power consumption of the UE is illustrated in Figure 33 for different devices operating in LTE band 7 (2.6 GHz).One can see from the plot that the power consumption curve can be divided into two parts: for a low transmission power, below a device specific threshold γ, the graph is characterized by a small, almost horizontal slope. For higher transmission power values, the power amplifier switches the mode and the slope becomes significantly steeper. This specific characteristic can be observed for all UEs under test.

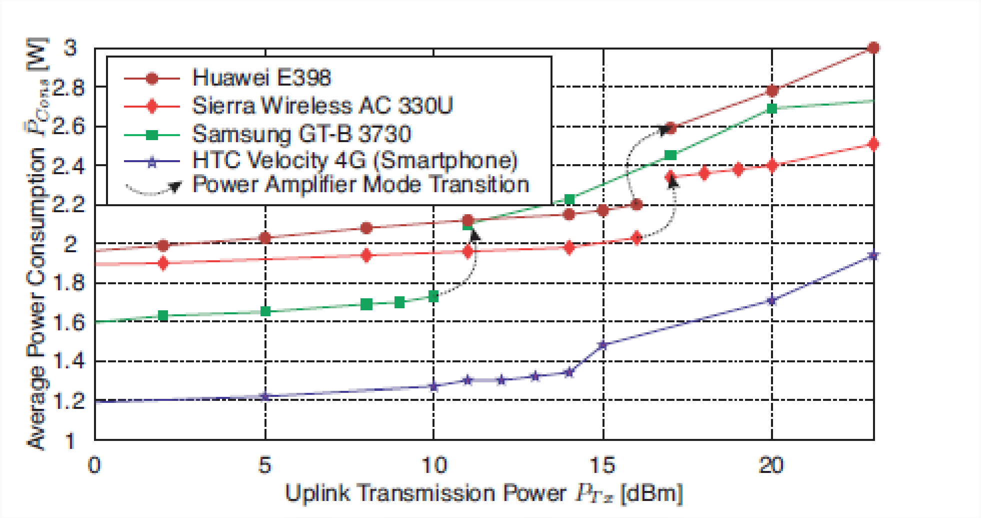


Figure 33: Device-specific Power Consumptions for Different Uplink Transmission Power Values (all in 2.6 GHz).

The power consumption curve can be approximated by to linear functions:

$$\bar{P}(P_{Tx}) = \begin{cases} \alpha_L \cdot P_{Tx} + \beta_L & \text{for } P_{Tx} \leq \gamma \\ \alpha_H \cdot P_{Tx} + \beta_H & \text{for } P_{Tx} > \gamma \end{cases}$$

Table 8: Model parameters for defferent LTE EU devices.

Model parameter	**HTC Velocity 4G**		**Samsung GT-B 3740**	**Samsung GT-B 3730**	**Huawei E 398**	**Sierra Wireless AC 330U**	
Frequency [MHz]	800	2600	800	2600	1800	2100	2600
α_L [mW/dBm]	4.8	4	7.7	7.2	10	5.6	5.4
β_L [W]	1.6	1.2	1.6	1.6	1.7	1.6	1.9
α_H [mW/dBm]	68	61	13	54	24	27	28
β_H [W]	0.79	0.52	0.4	1.5	1.9	1.5	1.8
γ [dBm]	12	12	11	10	16	16	16
$\dot{P}_{IDLE}$ [mW]	40	40	175	44	236	63	63
Maximum Error [%]	5.1	3.5	1.7	3.9	4.7	3.6	1.5

α and β are specific device parameters and y is the threshold that triggers the amplifier. The values are displayed in the Table 8, L and H stands for low power mode and high-power mode.

The table provides an approximation error for all devices. As an example, Figure 34 illustrates in detail the good match between the proposed model and the actual measurements for the HTC velocity 4G smartphone and both frequency bands supported by this device.

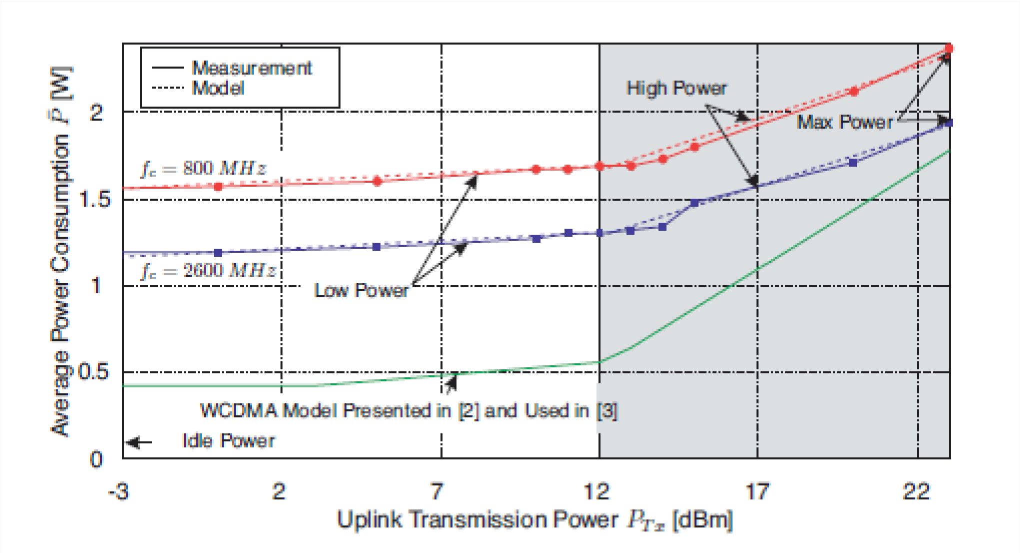


Figure 34: Tx-Power Dependent Average Power Consumption vs. Empirical Model.

4.4 Trade-Off Analysis of Research Solutions Summary

In this section the results of the analysis of the research solutions (Chapter 4.3) are presented, particularly, the comparison of the performance in terms of power consumption for WSN layer 2 and the comparison between security features in diverse WSN technologies.

This is presented with the aim to cover the two most important aspects of the future trackside solution: low energy consumption and high safety and security requirements.

4.4.1 WSN Layer 2 Trade-off

Table 9: Comparison of Line-MAC, S-MAC and Area-MAC protocols.

Protocol	**Topology**	**Consumed energy in linear topology**	**Relay burden problem**	**PDR in N1 best cases**	**PDR N1N2 best cases**	**Wake up /sleep techniques**
Line-MAC	Linear	Between 5J and 10J seconds depending on IT	No	91.88% in 10s of IT 91.86% in 15s of IT	94.04% in 5 s of IT 94.14% in 7s of IT	Yes
S-MAC	Grid, Mesh,	12,5 Joules	Yes	50.47% in 15	62.72% in 15 s	Yes

	Linear, Tree.			s of IT 62.76% in 20 s of IT	of IT 89.16% in 20s of IT	
Area-MAC	Grid, Mesh, Linear, Tree.	35 Joules	Yes	45% aprox in node n64	63% aprox in node n64	Yes

Table 9 provides a comparison between protocols adequate for LWSN considering the various characteristics previously outlined in Chapter 4.3.

4.4.2 Security Implementations for WSN

Table 10: Comparison of security features.

Type	Encryption	Block cipher	Freshness	Code Requirements	Auth provided	Cost (time/energy)	Key Agreement	Year
SPINS	Yes - CTR mode	RC5	Yes	2674 Bytes Max	Yes – CBCMAC	7.2ms/ 20 %	Master Key & Delayed Disclosure	2002
TinySec	Optional - CBC mode	Skipjack	NO	7146 Bytes Max	Yes – CBCMAC	0.38 ms/ 9.1%	Any	2004
LEAP	Yes - RC5	RC5	NO	17.8 Kb	Yes – CBCMAC	Variable (No. of neighbours)	Pre-deployed (Master) Variable	2003 (Implemented 2006)
SM	Yes - ECC	N/A	NO	N/A	EC-MQV	Variable (Nodes, parameters etc.)	EC-MQV Initial trust	2006
ZigBee	Optional - AES	AES-128	Yes ccm	N/A	Yes – CBCMAC	Under Investigation - Expected increase in latency and power consumption	SKA Trust Centre	2005

Table 10 above illustrates the various characteristics of the previously discussed security architectures under common headings. As it is evident, Symmetric key cryptography-based architectures have been the main source of security in Wireless Sensor Networking to date. There is much research available claiming that Public Key based solutions will provide better solutions, based on smaller key sizes and less storage requirements (under ECC), for more secure communications, also even providing superior energy efficiency. From an authentication perspective, the CBC-MAC algorithm is the most popular method of providing authentication for symmetric key based algorithms. Table above illustrates that it is chosen in all implementations of security architectures for WSNs to date. From a design perspective, scalability of security architectures is a desirable feature. Not all applications of WSNs will require the same security, and even in applications that do, different types of messages will require different degrees of security

4.4.3 LTE Power Consumption

Table 11: LTE Power consumption for different Modulation Coding Schemes.

Modulation Coding Scheme	Constellation	Resource Blocks	Signal To Noise Ratio for a Block To Error Ratio < 1% (db)	Power Consumption per RB	Power saving
MCS 17	16-QAM	1	14.4	5,012 mW	NA
MCS 11	16-QAM	2	8.6	1,259 mW	23.8%
MCS 9	QSPK	4	4.7	0,501 mW	25.7 %
MCS 2	QPSK	8	2.8	0,316 mW	27.1%
MCS 0	QPSK	12	1.6	0,251 mW	27.5%

5. CONCLUSIONS

In this document the current solutions and the State of the Art in research have been outlined for both On-board and Track-side communication.

The analysis has taken under consideration many factors, such as energy consumption, reliability of the network and employed technology for both hardware and software. After that, these solutions have been further examined and compared from the ETALON application perspective.

It is worth remarking that the analysed solutions do not employ energy harvesting technologies while this is a key aspect of the ETALON project. Furthermore, the analysed solutions have not taken into account safety aspect while the ETALON system should must be SIL4 “able”.

Solutions converge to the usage of WSN, or derivate from such, for monitoring train parameters and Train Integrity. The networks that have been analysed base the RF communication on different frequencies, finally resulting in the most reliable being a sub-GHz RF communication.

Regarding the On-Board, the solutions use different approaches for the generation of the network. When mounted on the train, each node does not know to which WSN it must connect in order to form the correct train network, introducing the problem of network awareness. The outlined solutions propose approaches such as measuring acceleration, node registration or distance awareness.

6. REFERENCES

[1] NGTC Grant Agreement, Annex I - “Description of Work”, 16th of August 2013.

[2] NGTC Consortium Agreement, V2, 12th of July 2013.

[3] Trainspotting, a WSN-based Train Integrity System, University of Twenty, Netherlands, 2009.

[4] Sensing Train Integrity, University of Twenty, Netherlands, 2009.

[5] An Innovative method of Train Integrity Monitoring Through Wireless Sensor Network, University of Florence, Italy, 2015.

[6] Railway Hazardous Articles Monitoring System Based on Wireless Sensor Network, Beijin Jiaotong University, China, 2010.

[7] Reliability Experiments for Wireless Sensor Networks in Train Environment, Uppsala University, Sweden, 2009.

[8] Linear Sensor Networks: Applications, Issues and Major Research Trends, Swaroop ISM, Dhanbad ISM, Dhanbad Galgotia University, India, 2015.

[9] Linear Wireless Sensor Network in M2M Communications: MAC layer Protocol Comparison, Ege University, Izmir, Turkey, 2016.

[10] Investigating energy efficiency and timeliness for linear wireless sensor networks, Radosveta Sokullu, Eren Demir, Ege University, Department of Electrical and Electronics Engineering, İzmir 35040, Turkey, 2014.

[11] A Distributed Topology Discovery Algorithm for Linear Sensor Networks, Imad Jawhar, Nader Mohamed and Liren Zhang, Faculty of Information Technology, UAE University, Alain, UAE, 2012.

[12] Modelling the Performance of Faulty Linear Wireless Sensor Networks; Nader Mohamed, Jameela Al-Jaroodi, and Imad Jawhar, The College of Information Technology, UAE University, P.O. Box 15551, Al Ain, UAE, University of Pittsburgh, Pittsburgh, PA 15260, USA, 2014.

[13] X2R-WP07-D-TTS-001-01-D7.1 Analysis of existing lines and economic models. X2Rail-1 deliverable, 2017

[14] X2R-WP03-D-SIE-005-01-D3.1 User & System Requirements (Telecommunications). X2Rail-1 deliverable, 2017

[15] Securing Wireless Sensor Networks: Security Architectures; David Boyle, Thomas Newe, Department of Electronic and Computer Engineering, University of Limerick, Limerick, Ireland, 2008.

[16] Utilizing Unused Network Capacity for Battery Lifetime Extension of LTE Devices BJoern Dusza, Christoph Ide and Christian Wietfeld from the Communication Networks Institute of Dortmund University, 2012.

[17] An Accurate Measurement-Based Power Consumption Model for LTE Uplink Transmission. 2013 IEEE Conference on Computer Communications Workshops, 2013.

[18] Effect of relay location on two-way DF and AF relay for multi-user system in LTE-A cellular networks Jaafar. A. Aldhaibani; A. Yahya; R. B. Ahmad; Normaliza Omar; Zaid G. Ali. 2013 IEEE Business Engineering and Industrial Applications Colloquium (BEIAC), 2013.

[19] White Paper Telesystems innovations – LTE in a nutshell: The physical layer, 2010.

[20] Regulatory status for using RFID in the UHF spectrum - link

[21] FFFIS for Eurobalises - link