



N° d'ordre NNT : 2022ISAL0120

**THÈSE de DOCTORAT DE L'INSA LYON,
membre de l'Université de Lyon**

**École Doctorale N° 512
(Informatique et Mathématique)**

Spécialité : Informatique

Soutenué publiquement le 16/12/2022, par :

Safuriyawu Ahmed

**Resilient IoT-based Monitoring System for the
Nigerian Oil and Gas Industry**

Devant le jury composé de :

Menaud, Jean-Marc
Stolf, Patricia
Guidec, Frédéric
Caron, Eddy
Takruri-Rizk, Haifa
Silva, Bhagya Nathali

Professeur des Universités
Professeur des Universités
Professeur des Universités
Maître de Conférences
Professeur des Universités
Senior Lecturer

Le Mouël, Frédéric
Stouls, Nicolas
Yusuf, Kabir

Professeur des Universités
Maître de Conférences
Docteur en gestion
des ressources
environnementales

IMT-Atlantique
Université de Toulouse
Université Bretagne Sud
ENS Lyon
University of Salford
University of Sri
Jayawardenepura
INSA Lyon
INSA Lyon
PTDF/SAPZ

Président
Rapporteuse
Rapporteur
Examineur
Examinatrice
Examinatrice

Directeur de thèse
Co-Directeur de thèse
Invité

Département FEDORA – INSA Lyon - Ecoles Doctorales

SIGLE	ECOLE DOCTORALE	NOM ET COORDONNEES DU RESPONSABLE
CHIMIE	<u>CHIMIE DE LYON</u> https://www.edchimie-lyon.fr Sec. : Renée EL MELHEM Bât. Blaise PASCAL, 3e étage secretariat@edchimie-lyon.fr	M. Stéphane DANIELE C2P2-CPE LYON-UMR 5265 Bâtiment F308, BP 2077 43 Boulevard du 11 novembre 1918 69616 Villeurbanne directeur@edchimie-lyon.fr
E.E.A.	<u>ÉLECTRONIQUE, ÉLECTROTECHNIQUE, AUTOMATIQUE</u> https://edeea.universite-lyon.fr Sec. : Stéphanie CAUVIN Bâtiment Direction INSA Lyon Tél : 04.72.43.71.70 secretariat.edeea@insa-lyon.fr	M. Philippe DELACHARTRE INSA LYON Laboratoire CREATIS Bâtiment Blaise Pascal, 7 avenue Jean Capelle 69621 Villeurbanne CEDEX Tél : 04.72.43.88.63 philippe.delachartre@insa-lyon.fr
E2M2	<u>ÉVOLUTION, ÉCOSYSTÈME, MICROBIOLOGIE, MODÉLISATION</u> http://e2m2.universite-lyon.fr Sec. : Sylvie ROBERJOT Bât. Atrium, UCB Lyon 1 Tél : 04.72.44.83.62 secretariat.e2m2@univ-lyon1.fr	M. Philippe NORMAND Université Claude Bernard Lyon 1 UMR 5557 Lab. d'Ecologie Microbienne Bâtiment Mendel 43, boulevard du 11 Novembre 1918 69 622 Villeurbanne CEDEX philippe.normand@univ-lyon1.fr
EDISS	<u>INTERDISCIPLINAIRE SCIENCES-SANTÉ</u> http://ediss.universite-lyon.fr Sec. : Sylvie ROBERJOT Bât. Atrium, UCB Lyon 1 Tél : 04.72.44.83.62 secretariat.ediss@univ-lyon1.fr	Mme Sylvie RICARD-BLUM Institut de Chimie et Biochimie Moléculaires et Supramoléculaires (ICBMS) - UMR 5246 CNRS - Université Lyon 1 Bâtiment Raulin - 2ème étage Nord 43 Boulevard du 11 novembre 1918 69622 Villeurbanne Cedex Tél : +33(0)4 72 44 82 32 sylvie.ricard-blum@univ-lyon1.fr
INFOMATHS	<u>INFORMATIQUE ET MATHÉMATIQUES</u> http://edinfomaths.universite-lyon.fr Sec. : Renée EL MELHEM Bât. Blaise PASCAL, 3e étage Tél : 04.72.43.80.46 infomaths@univ-lyon1.fr	M. Hamamache KHEDDOUCI Université Claude Bernard Lyon 1 Bât. Nautibus 43, Boulevard du 11 novembre 1918 69 622 Villeurbanne Cedex France Tél : 04.72.44.83.69 hamamache.kheddouci@univ-lyon1.fr
Matériaux	<u>MATÉRIAUX DE LYON</u> http://ed34.universite-lyon.fr Sec. : Yann DE ORDENANA Tél : 04.72.18.62.44 yann.de-ordenana@ec-lyon.fr	M. Stéphane BENAYOUN Ecole Centrale de Lyon Laboratoire LTDS 36 avenue Guy de Collongue 69134 Ecully CEDEX Tél : 04.72.18.64.37 stephane.benayoun@ec-lyon.fr
MEGA	<u>MÉCANIQUE, ÉNERGÉTIQUE, GÉNIE CIVIL, ACOUSTIQUE</u> http://edmega.universite-lyon.fr Sec. : Stéphanie CAUVIN Tél : 04.72.43.71.70 Bâtiment Direction INSA Lyon mega@insa-lyon.fr	M. Jocelyn BONJOUR INSA Lyon Laboratoire CETHIL Bâtiment Sadi-Carnot 9, rue de la Physique 69621 Villeurbanne CEDEX jocelyn.bonjour@insa-lyon.fr
ScSo	<u>ScSo*</u> https://edsciencessociales.universite-lyon.fr Sec. : Mélina FAVETON INSA : J.Y. TOUSSAINT Tél : 04.78.69.77.79 melina.faveton@univ-lyon2.fr	M. Christian MONTES Université Lumière Lyon 2 86 Rue Pasteur 69365 Lyon CEDEX 07 christian.montes@univ-lyon2.fr

*ScSo : Histoire, Géographie, Aménagement, Urbanisme, Archéologie, Science politique, Sociologie, Anthropologie

Abstract

Pipeline failures in crude oil transportation occur due to ageing infrastructure, third-party interferences, equipment defects and naturally occurring failures. Consequently, hydrocarbons are released into the environment resulting in environmental pollution, ecological degradation, and unprecedented loss of lives and revenue. Hence, multiple leakage detection and monitoring systems (LDMS) are employed to mitigate such failures. More recently, these LDMS include Wireless Sensor Networks (WSN) and Internet of Things (IoT)-based systems. While they are proven more efficient than other LDMS, many challenges exist in their adoption for pipeline monitoring. These include fault tolerance, energy consumption, accuracy in leakage detection and localisation, and high false alarms, to cite a few.

Therefore, our work seeks to address some of these challenges in implementing IoT-based systems for crude oil pipelines in a resilient and end-to-end manner. Specifically, we consider the aspect of accurate leakage detection and localisation by introducing a unique node placement strategy based on fluid propagation for sensitive and multi-sized leakage detection. We also propose a new distributed leakage detection technique (HyDiLLEch) in the WSN layer based on a fusion of existing leakage detection techniques such as the negative pressure wave method, gradient-based method, and pressure point analysis. With HyDiLLEch, we efficiently eliminate single points of failure associated with classical centralised systems.

Furthermore, we implement fault-tolerant data and service management in the fog layer utilising the Nigerian National Petroleum Corporation (NNPC) pipeline network as a use case. The problem is modelled as a regionalised data-driven game against nature on the NNPC pipelines. Our proposed regionalised solution (R-MDP) using reinforcement learning optimises accuracy and fault tolerance while minimising energy consumption.

Overall, our system guarantees resiliency to failures and efficiency in terms of detection and localisation accuracy and energy consumption.

Résumé

Les défaillances d'oléoducs dans le transport du pétrole brut se produisent en raison du vieillissement de l'infrastructure, des interférences de tiers, des défauts d'équipement et des défaillances naturelles. Par conséquent, des hydrocarbures sont rejetés dans l'environnement, entraînant une pollution de l'environnement, une dégradation écologique et des pertes de vies et de revenus sans précédent. Par conséquent, plusieurs systèmes de détection et de surveillance des fuites (LDMS) sont utilisés pour atténuer ces défaillances. Plus récemment, ces LDMS incluent les réseaux de capteurs sans fil (WSN) et les systèmes basés sur l'Internet des objets (IoT). Bien qu'ils se soient avérés plus efficaces que d'autres LDMS, de nombreux défis existent dans l'adoption de tels systèmes pour la surveillance des pipelines. Ceux-ci incluent la tolérance aux pannes, la consommation d'énergie, la précision de la détection et de la localisation des fuites et le nombre élevé de fausses alarmes, pour n'en citer que quelques-uns.

Par conséquent, notre travail vise à relever certains défis dans la mise en œuvre de systèmes basés sur l'IdO pour les oléoducs de pétrole brut de bout en bout de manière résiliente. Plus précisément, nous considérons les aspects de détection et localisation précises des fuites en introduisant une stratégie de placement de nœud unique basée sur la propagation des fluides pour une détection de fuite sensible et multi-tailles. Nous proposons également une nouvelle technique de détection de fuite distribuée (HyDiLLEch) dans la couche WSN basée sur une fusion des techniques de détection de fuites existantes telles que la méthode des ondes de pression négative, la méthode basée sur le gradient et l'analyse des points de pression. Avec HyDiLLEch, nous éliminons efficacement les points de défaillance uniques associés aux systèmes centralisés classiques.

En outre, nous mettons en œuvre une gestion des données et des services tolérante aux pannes dans la couche de infrastructure Edge en utilisant le réseau de oléoducs de la Nigerian National Petroleum Corporation (NNPC) comme cas d'utilisation. Le problème est modélisé par la théorie des jeux avec une approche régionalisée du réseaux NNPC contre la nature. Notre proposition de solution régionalisée (R-MDP) utilise l'apprentissage par renforcement et optimise la précision et la tolérance aux pannes tout en minimisant la consommation d'énergie.

Dans l'ensemble, notre système garantit la résilience aux pannes et l'efficacité en termes de précision de détection et de localisation et de consommation d'énergie.

Dedication

To my late father, Sheikh Ahmed Rufai AbdulKareem, I bear your name with a deep sense of pride and I hope you are proud of me wherever you are. May Allah SWT grant you Jannatul Firdaus, Papa.

To my queen mother of inestimable value, around whom my world revolves, Hajia Fatimah Ahmed Rufai, you're the definition of strength, love, perseverance, and patience. Onyam oziorijo, my role model, I am forever grateful for your constant prayers and support.

To my beloved brother, without whom I will not be where I am today, Engr. Timasaniyu Ahmed Rufai (ohinoyi onoru), I owe this to you. Thank you for being a girl's hero, I am deeply grateful for all you have done for me.

To my sisters: (Hajia) Mabrukat, Rabiyyatu Al-Adawiya, Hajara, and Bilqis, my league of powerful and extraordinary women, enyene okumi okumi, I thank you for laying a solid foundation for me. On your shoulders, I rose, stood tall, and reached places that would have been otherwise impossible, avo nini. I immensely appreciate you for holding my hands through the journey of life. As I continue towards the future, I know I can always rely on your unceasing support, love, and prayers.

This work is hereby dedicated to you ALL.

Declaration

I declare that the contents of this dissertation are original except for the references made to other research works. It is written solely from the results of the work I conducted myself and has not been submitted for consideration for any other degree or qualification in this, or any other university. It is written based on the following works either published or accepted for publication:

1. S. Ahmed, Frédéric Le Mouël, and N. Stouls, “Resilient IoT-based Monitoring System for Crude Oil Pipelines,” in Proceedings of the 7th International Conference on Internet of Things: Systems, Management and Security (IOTSMS). IEEE, Dec. 2020.
2. S. Ahmed, Frédéric Le Mouël, N. Stouls, and G. L. Kouyi, “HyDiLLEch: a WSN-based Distributed Leak Detection and Localisation in Crude Oil Pipelines,” in Proceedings of the 35th International Conference on Advanced Information Networking and Applications (AINA-2021). Springer, May 2021. (Rank A2/B)
3. S. Ahmed, F. Le Mouël, N. Stouls, and Jilles Dibangoye, “R-MDP: A Game Theory Approach for Fault-Tolerant Data and Service Management in Crude Oil Pipeline Monitoring System,” in Proceedings of the 19th International Conference on Mobile and Ubiquitous Systems (MobiQuitous). Springer, Nov. 2022. (Rank A/B1)

Acknowledgements

I want to first thank Almighty Allah (SWT) for his mercy and grace upon my life.

To Professor Frédéric Le Mouël: I am profoundly thankful for your undiluted interest, contributions and commitment to this work. You mentored me and gave me a platform for personal and professional growth and your incomparable vision saw this work to conclusion, merci beaucoup.

To Assoc. Professor Nicolas Stouls: Your contribution to this work is immense and words cannot express how grateful I am. I thank you also for your patience, trust, time, and most importantly your dedication during the course of this work.

I would also like to thank Assoc. Prof Gislain Lipeme Kouyi and Assoc. Prof Jilles Steven Dibangoye for their valuable contributions to the hydraulics and game theory parts of this work, respectively.

To my family members and friends who have supported me in one way or the other in the course of this journey, I thank you all. I pray that Allah SWT continue to keep us together.

Finally, I am especially grateful to the Petroleum Technology Development Fund (PTDF) for financing this work.

Contents

I	BACKGROUND	2
1	Introduction	3
1.1	Context	3
1.1.1	Crude oil transportation	4
1.1.2	Pipeline failures and causes	6
1.1.3	Effects of pipeline failures	9
1.1.4	Summary	12
1.2	Problem Definition	12
1.3	Contributions	16
1.3.1	Wireless Sensor Network Layer	16
1.3.2	Data and Service Layers	16
1.4	Thesis Organisation	17
1.4.1	Part One: Background	17
1.4.2	Part Two: Contributions	17
2	State-of-the-Art	19
2.1	Classical monitoring of pipeline transportation of crude oil	19
2.1.1	Sensors for pipeline monitoring	20
2.1.2	Fluid Mechanics	24
2.1.3	Pipeline Failure Detection and Localisation Techniques	30
2.1.4	Summary and Conclusion	32
2.2	Industrial IoT and its application to pipeline monitoring	35
2.2.1	Architectural Design of IoT Systems	36
2.2.2	Wireless Sensor Placement Strategies	37
2.2.3	WSN-based pipeline monitoring	38
2.2.4	Data and Service Placement	40
2.2.5	Processing, learning and analytics	42
2.2.6	Summary and Conclusion	45
II	CONTRIBUTIONS	47
3	System Design and Specifications	48
3.1	Background	48
3.1.1	IoT Generic and Application-based Architectures	48
3.1.2	Wireless Communication Protocols for IoT	50

3.2	Design Metrics and Considerations	52
3.2.1	Reliability and robustness	53
3.2.2	Coverage and connectivity	54
3.2.3	Accuracy of leakage detection and localisation	54
3.2.4	Energy consumption	54
3.3	System design and discussions	55
3.3.1	Architectural design	55
3.3.2	Communication protocols	56
3.3.3	Event coverage	58
3.3.4	Data management	59
4	Distributed Detection and Localisation of Leakages in Pipelines	61
4.1	Collective Detection of Multifarious leakages	62
4.1.1	Node Placement	62
4.1.2	Spatial Data Correlation	64
4.2	HYbrid DIstributed Leakage detection and Localisation tEChnique (HyDiLLEch)	65
4.2.1	3-factor Leakage Detection	68
4.2.2	2-factors Leakage Localisation	70
4.3	Simulation Results	71
4.3.1	Node placement	71
4.3.2	Detection and Localisation	73
4.3.3	Communication Efficiency	76
4.4	Conclusion	80
5	A Game Theory Approach to Data and Service Management for Crude Oil Pipelines Monitoring Systems	82
5.1	Failures in pipeline transportation and modelling process	83
5.1.1	Introduction to the Environment and Problem Definition	83
5.1.2	The Environment Setting	87
5.2	R-MDP: A Data-Driven Regionalised Markov Decision Process	88
5.2.1	The First Stage: Accuracy Optimisation	88
5.2.2	The Second Stage: Optimising the Energy Consumption	91
5.3	Implementation and Results	93
5.3.1	Accuracy in Detection and Fault Tolerance	96
5.3.2	Optimised Accuracy and Energy Consumption	99
5.4	Conclusion and Discussion	103

6	Conclusion and Future Works	104
6.1	Summary	104
6.2	Future Works	105

List of Figures

1.1	The three sectors of the Oil and Gas Industry	4
1.2	Crude oil transportation by mode in the US [1]	6
1.3	Statistical analysis of causes of pipeline failures in liquid and gas pipelines in the US [2]	7
1.4	Crusted land and swampland pollution [3]	10
1.5	Hydrocarbon pollution of water bodies [3]	11
1.6	Digital maturity of the midstream sector [4]	14
1.7	Forecasted growth of connected devices	15
2.1	Steady fluid transmission in a pipeline	24
2.2	Effects of leakage on the pressure gradient and NPW generation . . .	25
2.3	Pressure points measurements in a pipeline	26
2.4	The pressure gradient distribution after a leak	27
2.5	Negative pressure wave introduced by a leakage	28
3.1	Generic IoT Architecture [5]	49
3.2	Wireless Communication Technologies [6]	51
3.3	Network Architecture	56
3.4	Network Architecture with Communication	57
3.5	Event coverage on the proposed architecture	58
3.6	Data and Service Placement	60
4.1	Sensor placement on long transmission crude oil pipelines	63
4.2	Detection and localisation of leakages	68
4.3	Leakage detectability by number of sensors and distance	72
4.4	NPWM and GM	74
4.5	HyDiLLEch-1 Average localisation accuracy by NDLs	76
4.6	HyDiLLEch-2 Average localisation accuracy by NDLs	77
4.7	Communication overhead by the number of packets	78
4.8	Sampling energy consumption of the sensors	79
4.9	Radio energy consumption	79
4.10	Cumulative energy consumption	80
5.1	Nigerian National Petroleum Corporation pipeline network [7]	84
5.2	5 year regionalised pipeline incidents [8]	84
5.3	State transition Diagram	87
5.4	Simulated pipeline network and OpenAI Gym Interaction	94
5.5	Total Number of Nodes Detecting Leakage	97
5.6	Total Number of HPN	98

5.7	Accuracy in Leakage Detection from all nodes	98
5.8	Accuracy in Leakage Detection in HPN	99
5.9	Total Obtainable Reward across the Error Rates	100
5.10	Total Obtainable Rewards with HPN	100
5.11	Total Reward by Algorithm	101
5.12	Total Energy Consumption by Algorithm	102
5.13	Regional Reward vs Energy Consumption	102

List of Tables

1	List of acronyms	1
1.1	Causes of Pipeline Failures by the transportation mode	12
2.1	Summary of pipeline monitoring techniques	33
2.2	Summary of pipeline monitoring sensors and systems	35
2.3	Summary of common IoT challenges in pipeline monitoring and applied approaches	46
3.1	Properties of some IoT Wireless Communication Protocols	52
4.1	Pipeline and oil characteristics	73
4.2	Network simulation parameters	73
5.1	Simulation Parameters	94

Acronyms

3GPP	3rd Generation Partnership Project
AI	Artificial Intelligence
BFT	Byzantine Fault Tolerance
BLE	Bluetooth Low Energy
CCTV	Closed-Circuit Television
CDN	Content Distribution Network
CFD	Computational Fluid Dynamics
CPS	Cyber Physical Systems
DAL	Detection and Localisation
DBMS	Database Management Systems
DPR	Department of Petroleum Resources
dwt	dead-weight tonnage
GA	Greedy Approximation
GM	Gradient-based Method
GMM	Gaussian Mixture Model
GPR	Ground Penetrating Radars
GPS	Global Positioning System
GSM	Global System for Mobile Communication
GSP GRS	Greedy Service Placement with Greedy Request Scheduling
GSP ORS	Greedy Service Placement with Optimal Request Scheduling
HPN	High Performing Nodes
HyDiLLEch	Hybrid Distributed Leakage Detection and Localisation Technique
IaaS	Infrastructure as a Service
IEEE	Institute of Electrical and Electronics Engineers
IoT	Internet of Things
KNN	K-nearest Neighbour
LAS	Leakages and Spills
LDMS	Leak Detection and Monitoring System
LDT	Leakage Detection Technique
LNG	Liquefied Natural Gas
LoRaWAN	Long Range Wide Area Network)
LPWAN	Low Power Wide Area Network
MDP	Markov Decision Process
MEMS	Micro-electromechanical System
MIO	Mixed Integer Optimisation
MIP	Mixed Integer Programming
MTC	Minimum Test Cover

NB-IoT	Narrowband Internet of Things
NDL	Nodes Detecting and Localising Leakages
NNPC	Nigerian National Petroleum Corporation
NPW	Negative Pressure Wave
NPW_{ds}	Negative Pressure Wave Detecting Sensor
NPW_{mdd}	Negative Pressure Wave Maximum Detecting Distance
NPWM	Negative Pressure Wave Method
OGI	Oil and Gas Industry
ORS	Optimal Request Scheduling
Paas	Platform as a Service
PG	Pressure Gradient
PHMSA	Pipeline and Hazardous Materials Safety Administration
PIG	Pipeline Inspection Gauges
POD	Place of Deployment
PPA	Pressure Point Analysis
PSI	Pounds per Square Inch
pubsub	publish subscribe
QoS	Quality of Service
RFID	Radio Frequency Identification
RGA	Robust Greedy Approximation
RLHO	Reinforcement Learning Heuristic Optimisation
RMIO	Robust Mixed Integer Optimisation
RSFO	Robust Sub-modular Function Optimisation
SaaS	Software as a Service
SCADA	Supervisory Control and Data Acquisition
SPOF	Single Points of Failure
SVM	Support Vector Machine
UAV	Unmanned Aerial Vehicles
UDP	User Datagram Protocol
ULCC	Ultra Large Crude Carrier
UMTS	Universal Mobile Telecommunication System
UNEP	United Nations Environment Programme
USD	United States Dollars
VLCC	Very Large Crude Carrier
WSC	Weighted Set Cover
WSN	Wireless Sensor Networks

Table 1: List of acronyms

Part I

BACKGROUND

Chapter 1

Introduction

The Oil and Gas Industry (OGI) is a significant part of the global economic framework. It plays a crucial role in the energy market as the primary fuel source, generating annual revenue in trillions of dollars globally. It is also the source of petrochemical feedstock and asphalt and serves as the leading energy supply for pharmaceutical products and solvents. Currently, the total annual global consumption of oil is approximately 30 billions barrel with a forecast of approximately 53% increment in consumption by 2035 [9].

Still, numerous challenges exist across the production processes of these products, from multiple system failures, inefficient infrastructural monitoring, and under-utilisation of data. In particular, the transportation of oil and gas products could result in leakages and spills (LAS) due to some of these challenges. Crude oil LAS is the release of liquid hydrocarbons into the environment causing environmental pollution, loss of lives from resulting fire incidents and degradation of biodiversity, among other grave consequences.

As such, over the years, companies, operators and regulatory bodies employed a range of measures, services and technological solutions to alleviate the impact of such failures. However, records have shown a continuous increase in LAS's resulting consequences, especially in developing countries like Nigeria. This is due to the complexity of mitigating such problems.

Thus, this chapter introduces our work on the design and development of a resilient IoT-based Monitoring System for the Nigerian OGI. In the first section, we present the various sectors of the OGI, from the drilling processes to the end user consumption. We also discuss the challenges generally faced in the Industry. In section 1.2, we narrow our discussion to the focus of the thesis, which is the pipeline transportation of crude oil. We enumerate the current solutions adopted to this sector's failures and limitations. In section 1.3, we briefly introduce our contribution to the work. Finally, we outline the organisation of the thesis in section 1.4.

1.1 Context

The OGI has a highly complex and capital-intensive set of procedures, from exploration to marketing, as shown in Fig. 1.1. The set of processes can be broadly classified as **upstream**, **midstream**, and **downstream** sectors.

The *upstream sector* includes exploration or search of crude oil, natural gas, and

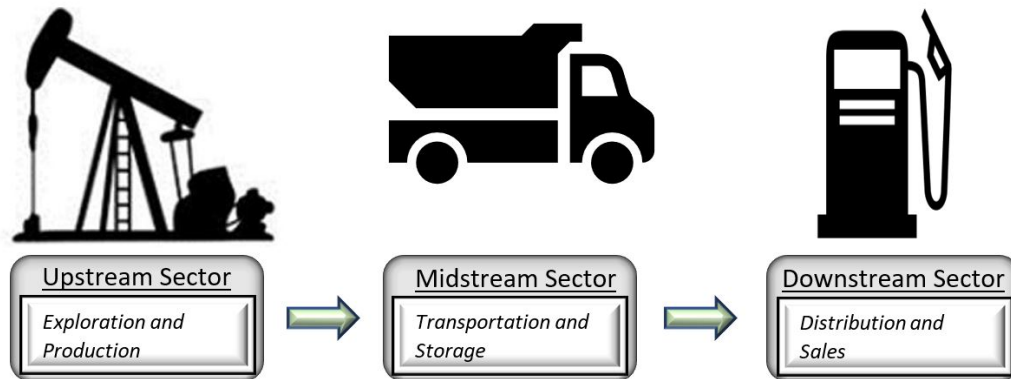


Figure 1.1: The three sectors of the Oil and Gas Industry

others in fields, underground or below the sea. When potential discoveries are made, the production process begins, which involves the drilling and operating of the discovery site. After production, the refined products are transported to the whole sellers, distributors and others in the supply chain. This part of the process is the *midstream sector*. Finally, the *downstream sector* concerns the purification, marketing and distribution of usable products, such as kerosene, gas, and diesel, to mention a few. Each sector has its challenges; for instance, the drilling processes in the upstream sector can be improved and accelerated using artificial intelligence (AI) or machine learning techniques [10, 11]. On the other hand, the supply chain in the downstream sector can be digitalised to gain valuable insights used in more efficient product distribution. This efficiency can be accomplished by the analysis of the big data resulting from the digitalisation of the sector [12, 13]. The midstream sector especially, lags in its digitalisation more than the other sectors, causing devastating effects.

Hence, we focus this research on the midstream sector. We worked on this sector specifically as it relates to the pipeline transportation of crude oil, attributed failures and solutions. The following subsections detail the midstream sector, incidents, and their environmental and socio-economic impacts.

1.1.1 Crude oil transportation

As aforementioned, several processes enable the transformation of oil and gas from raw to finished products. The midstream sector, i.e. transportation sector, is a vital part of the supply chain, facilitating these processes. For crude oil, the distribution is enabled through multiple means, such as pipelines, ships, trucks and rails [14–16] to the appropriate recipient in the supply chain. Following are brief details of each

mode of transportation.

Pipelines

Pipeline transportation consists of a network of pipelines in thousands of kilometres divided into two categories, i.e. natural gas pipelines and liquid pipelines. Crude oil are transported using the liquid pipelines, subdivided into *gathering*, *feeding* and *transmission* lines. The gathering lines are used for short-distance transportation of oil, usually from wells to processing tanks, refineries and others. On the other hand, the feeder lines connect the tanks and or processing facilities to the main lines known as the transmission lines. The transmission lines are used for long-distance transportation of crude at a national or international level. To enable such long-distance transportation, certain pressure must be maintained. This necessitates adding pumping stations spanning the pipeline network to boost the pressure. Generally, pipelines are considered the safest means of crude oil transportation.

Ships

Ships are the second most used mode of crude oil exportation to foreign countries. The type of ships expressly built for transporting oils in the sea is called tankers. Tankers are grouped by their size in terms of their dead-weight tonnage (dwt), which is determined by the canals and traits via which they travel, e.g. the Cape of Good Hope, the Strait of Malaca, the Strait of Dover, the Suez Canal, or the Panama Canal. For example, Ultra Large Crude Carriers (ULCCs) tankers range from 300 000 dwt to 500 000 dwt. Very Large Crude Carriers (VLCCs) tankers have a maximum size of 300 000 dwt, while the Suemax tankers and the Aframax tankers have maximum sizes of 200 000 dwt and 120 000 dwt, respectively. This method of crude oil transportation is the slowest compared to the other methods.

Trucks

Trucks are considered the least efficient mode of transporting crude oil based on the accrued cost and the carbon footprint [1,17]. Although the most accessible mode of transporting oil, trucks offer a limited storage capacity for carrying the transported product. Despite this constraint, it is the first choice where the distance between origin and destination is short, typically in the final stage of the midstream sector. However, good roads are critical to allow efficient truck transportation of crude oil. A study of the state of Colorado revealed the ratio of total accidents recorded results from truck transportation of oil and gas products per the population capital is 95% [18]. It also resulted in the most cause of deaths [19–21]

Rails

Cargo trains are also another form of transporting oil and gas with the aid of specialised tankers cars. The tankers used in this case differ from those used for ship transportation in capacity. Rail transports are possible using rail tracks, usually from oil wells to refineries. For large volumes of oil, multiple railcars are used. This method of transportation is preferable to trucks as accidents only happen when a train derailes. It is also considered the alternative to the pipeline mode of transportation when they are at total capacity, as the infrastructure's existence makes it flexible.

Statistical comparison of crude oil transportation methods

Among these modes of crude oil transportation, pipelines are the most widely used as they are considered the safest option and have the lowest carbon footprint.

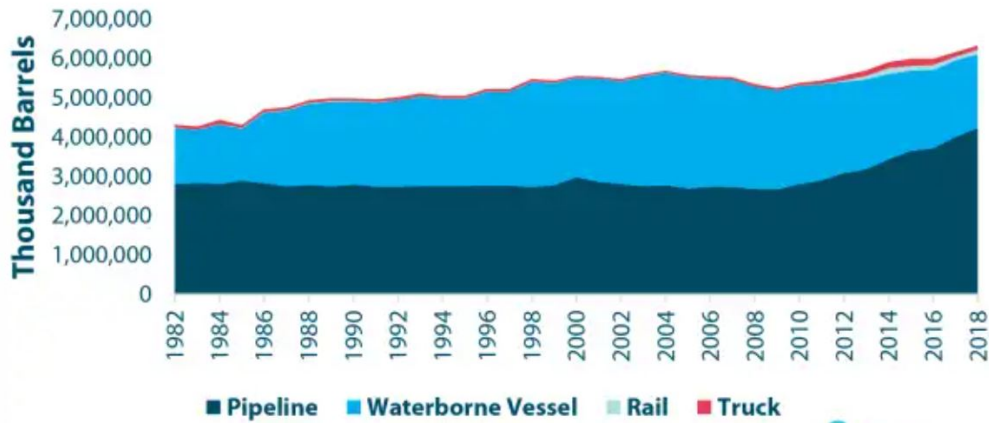


Figure 1.2: Crude oil transportation by mode in the US [1]

As seen in Fig. 1.2, pipelines have consistently been the choice for transporting crude oil and have shown an increase in recent years. Yet, a considerable amount of failures are still reported yearly [22]. Thus, in the following subsection, we discuss some common causes of pipeline failures and their impacts.

1.1.2 Pipeline failures and causes

According to regulatory bodies such as the Department of Petroleum Resources (DPR) of Nigeria and the United States Pipeline and Hazardous Materials Safety Administration (PHMSA) [2], causes of pipeline failures can be categorised as follows.

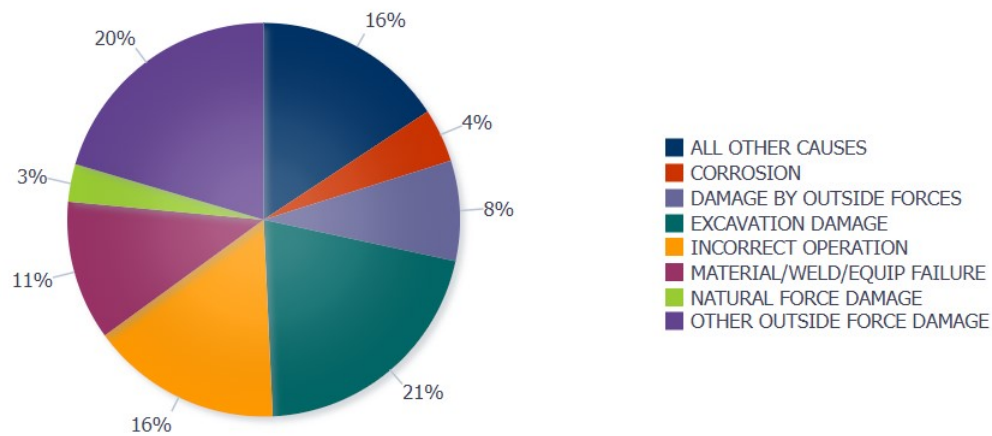


Figure 1.3: Statistical analysis of causes of pipeline failures in liquid and gas pipelines in the US [2]

Corrosion

Corrosion of a pipeline occurs due to a deteriorative natural process called oxidation with the pipeline's surroundings. Corrosion is time-dependent and results in metal loss or the wall thickness of the pipe leading to leakages if left uncontained for an extended period. It is of different types, i.e. internal, external, stress or microbes induced, stay current, selective and seam corrosion. Internal corrosion results from chemical reactions within the pipelines. External corrosion on the hand occurs as a result of environmental conditions on the external surface of the pipeline and the surroundings, such as water, soil and others. This type of failure accounts for about 18% of all causes of failures in liquid pipelines.

Natural Hazards

Natural events such as earthquakes, landslides, and extreme temperatures can subject pipelines to large-scale failures. While the mechanism is put in place to envisage some of these natural events, their unpredictability results in eventual pipeline failures. A one-year report between 2002 and 2003 shows that 9% of incidents resulted from natural hazards. Currently, operators are taking risk prevention steps through the identification, assessment and preparation for geotechnical and meteorological events.

Excavation

Excavation damage primarily concerns buried or underground pipelines. They occur due to various construction activities such as digging and trenching of roads, lands, and others. Consequently, pipelines can sustain dents, scrapes, punctures and others, resulting in future failures or damages causing immediate failures. These failures can be hazardous to the public due to the release of petroleum or gas products into the environment. About 15% of liquid pipeline failure incidents are caused by excavation.

Equipment Failure

While some failures occur directly on the pipeline, critical operational equipment such as pumps, compressors, meters, valves, sensors, and others can sometimes cause failures, inadvertently resulting in the spilling of hazardous fluid into the environment. However, the volume of products released to the environment due to this type of incident is small compared to others. Thus, equipment failure rarely causes significant harm to the public.

Material and Weld Failures

Material failures occur as a result of manufacturing defects, such as the oxidation of impurities resulting in what is known as lamination and inclusions. Blisters and scabs caused by a gas expansion in the pipeline material can also lead to material failures in pipelines. Weld failure, on the other hand, occurs from making or joining the pipes together during the initial construction phase or during maintenance. Weld failure in new pipelines includes pinholes, incomplete fusion, porosity and others. In older pipelines, weld failure includes weld metal cracks, hook cracks, and cold welds, among others.

Operational Failures

Operational failures pertain to the human factor aspect of pipeline failure, which could be company staff or contracted personnel. They occur due to the incorrect direction of the fluid, filling and draining of vessels or tanks, or carrying out routine maintenance of the pipelines. Although they are considered an indirect cause of failure, they still sometimes result in the release of harmful products to the environment. Operational failures, however, are not as prevalent as the other types of failures.

Third Party Interference

Third-party interference, otherwise known as "other outside force/damage by outside forces" i.e. sabotage, vandalism, or vehicle accidents, causes pipeline failure. The

frequency of such interferences depends on the location of the pipelines, i.e. there are more reported cases of vandalism in developing countries, e.g. Nigeria, than in the United States of America. In the presented data in Fig.1.3, this accounts for 28% of the causes of pipeline failure causes. Pipelines susceptible to third-party interference include overground pipelines. As we will discuss later in chapter 5, the proximity of the pipelines to highways and some communities also increases the rate of third-party interference.

1.1.3 Effects of pipeline failures

Regardless of the cause, pipeline failures have severe negative environmental and financial impacts. The effects of these failures largely depend on the pipeline type and location. For example, failures in overground pipelines could negatively affect the environment through land contamination. Likewise, failures in subsea pipelines could harmfully affect water bodies and aquatic life. LAS also results in fatalities from direct or indirect consequences and the livelihood of the communities co-located with the pipelines. Over the past 20 years, the US has recorded more than 5780 pipeline incidents, with approximately 0,045% fatalities and 19% injuries. These incidents resulted in costs and litigations to the tune of over 11.17 billion USD [2]. Thus, in this subsection, we discuss the impacts of pipeline failures in two broad categories: land water and biodiversity and its public health and socio-economic impact.

Land, Water and Biodiversity

Crude oil leakages and spills (LAS) cause contamination in lands, sediments, swamp-land, seas and oceans, altering the physical and chemical properties of the contaminated areas. Figures 1.4 and 1.5 are an example of the effect of spills on land, swampland and mangrove water bodies. The severity of the contamination depends on several factors. For example, when spillage or leaks occur on land, the composition of the type of soil on the land impacts its severity. In addition, the chemical composition of the spilt hydrocarbon and the spilt volume affects the severity of the impact. Depending on this volume, studies have shown that land contamination occurs deeper than the surface, i.e. contamination can go up to 5m into the ground. Undersurface contamination results in well water pollution.

Furthermore, fire accidents resulting in the surface crust on lands, as shown in Fig. 1.4, negatively affects the herbage of the land and can make re-vegetation of the land a herculean task or impossible in some cases. For example, the United Nations Environment Programme (UNEP) [3] evaluated LAS impact on a 122 km pipeline right of way in about 200 locations in Ogoniland of, Nigeria. According to

the environmental assessment, UNEP concluded that complete land restoration might take up to 30 years due to the extent of damage caused. Similarly, the life cycle of the vegetation is compromised by the loss of mineral nutrients and the gaseous exchange of the plants. Stress is also induced due to the disruption of ions, excluding capabilities from the roots of the plants.



Figure 1.4: Crusted land and swampland pollution [3]

In addition, spillage on water bodies, sea and mangroves occurs directly when a subsea pipeline fails. It may also indirectly impact surface and subsurface flows, rain, and wind affecting groundwater aquifers. Likewise, the water body's physical and chemical composition are also altered depending on the concentration level of the LAS. For example, oxygen transfer in the water column is prevented, resulting in the death of various plants such as pneumatophores, lenticels, trees, and seedlings. Additionally, this oxygen deprivation can partially or completely destroy the mangroves and unsettle the aquatic ecosystem. In Bonny local government in the Niger Delta region of Nigeria, over 307,380 square meters of healthy mangrove was lost due to artisanal refining, and the resulting LAS between 2007 and 2011 [3]. According

to the same report, the highest concentration of hydrocarbon found in underground water is approximately one million micrograms per litre, nearly 99.94% more than the recommended standard by the Nigerian government.

Public Health and Socio-economic Impacts

Leakages and spills cause fire accidents and explosions that have devastating effects, such as the deaths of thousands of people, grave body injuries, and respiratory problems, to list a few. Non-fire but other life-threatening effects have also been reported by various bodies [2, 3, 23]. In many cases, there is a notable reduction in the life expectancy of people in the host communities. For instance, this can be a direct consequence of the elevated hydrocarbon concentration in the air and water.

Also heightening the risk of sickness and possibly death is the exposure and consumption of contaminated water. These result in problems to the digestive, immune, respiratory and nervous systems and rashes through dermal contact. Besides, in the Ogoniland case study, benzene, a known cancer-causing hydrocarbon, was present in the air at a rate of 900 times more than the World Health Recommendation. Such elevated levels of benzene in the environment possess a great risk to the public and put them at risk of acute myeloid leukaemia, acute and chronic lymphocytic leukaemia, multiple myeloma, and non-Hodgkin lymphoma [24, 25].



Figure 1.5: Hydrocarbon pollution of water bodies [3]

Although the effect of LAS can sometimes be direct in terms of fatality, injuries or health-related issues, its effects on lands and sea also have direct socio-economic impacts at both local and national levels. For example, comparative studies [23] show that LAS-affected lands produce lower yields than non-affected lands. In some cases, it renders root crops unusable due to the high concentration of harmful hydrocarbons in the crops. Thus, farmers whose livelihood depends on such lands either lose their income or have to look for additional or alternative sources of income. Additionally, increments in school dropouts of children were recorded as an indirect consequence

of LAS.

So far, we have introduced the processes in OGI, particularly the midstream sector. We also discussed the causes and effects of failures in the midstream sector. In the next section, we define the problem addressed in this thesis as related to the pipeline transportation of crude oil.

1.1.4 Summary

In this subsection, we summarise the failure caused by the mode of transportation in the midstream sector of the OGI.

Mode / Cause	Corrosion	Regulatory Oversight	Natural Disasters	Collisions	Muffler Defects	Excavation	Third Party Interference	Human Error	Maintenance	Operational	Undetermined	Allisions	Grounding	Location	Monitoring
Pipelines	✓	✓	✓	•	•	✓	✓	✓	✓	✓	✓	•	•	✓	✓
Rails	•	✓	•	✓	•	•	•	✓	•	•	✓	•	•	•	•
Ships	•	✓	•	✓	•	•	•	✓	•	✓	✓	✓	✓	•	•
Trucks	•	•	•	✓	✓	•	✓	✓	•	•	✓	•	•	•	✓

Table 1.1: Causes of Pipeline Failures by the transportation mode

1.2 Problem Definition

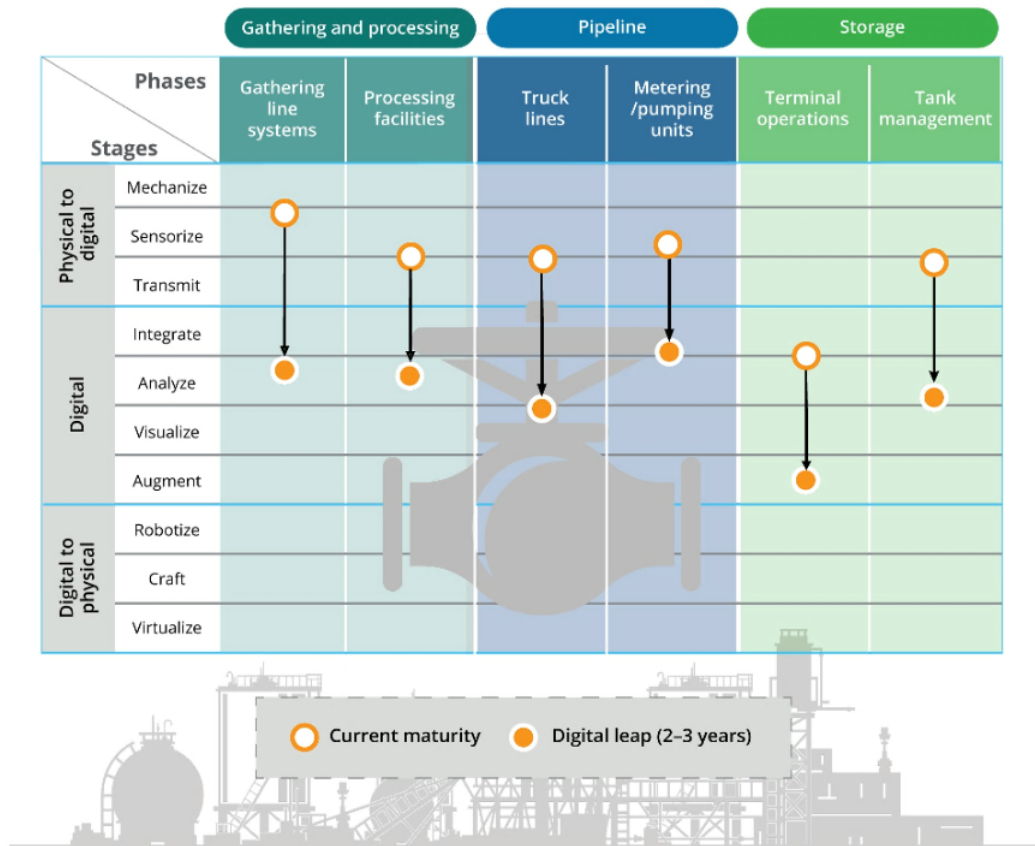
Pipelines, as introduced earlier, are critical national and international infrastructures used to transport fluids such as crude oil, natural gas, bitumen, or water. In the midstream sector, they are considered the safest and mostly used transportation means for oil and gas products [19], [26]. Despite its high safety rate, pipeline transportation can sometimes fail due to third-party interferences, equipment failures, corrosion and other causes of failures discussed in subsection 1.1.2 with vandalisation or third-party interference as the leading cause of failures in developing countries like Nigeria.

These failures led to an annual loss of approximately 10 billion USD in the United States [27] and 100 million USD in Nigeria, excluding the payments of litigations, fines, compensation and so on [19]. Furthermore, anomalous events such as leakages or sabotage have disastrous effects not just economically but also environmentally [28], [26]. LAS has resulted in fatal accidents, one of which recorded deaths of over a thousand in Jesse in 1998 [19]. Another pipeline incident in January 2019 was reported to have killed about 135 people in Mexico. More consequences of LAS have previously been discussed in subsection 1.1.3.

Therefore, it is paramount to have a system in place to monitor, in real time, the various operations of pipelines to detect any anomaly that occurs promptly. Consequently, some monitoring systems for oil and gas pipelines were employed. They include:

1. Wired systems using fibre optic cables or copper
2. Wireless Systems
3. A hybrid of both wired and wireless systems
4. Daily overflights using state-of-the-art high-definition cameras to a specialised helicopter
5. Surveillance by security personnel
6. Community-based surveillance
7. Unmanned aerial vehicles (UAVs), i.e.Drones
8. Robots

Despite implementing some of the techniques listed above for pipeline faults detection and monitoring, some countries like Nigeria have witnessed increasing incidents in their pipeline networks. In 2018, Shell Nigeria recorded a loss of petroleum of up to 11000 barrels per day in their pipeline network [29]. This loss was an increment of almost 550% compared to the reported loss in the previous year. In addition, some of the listed monitoring techniques, such as community-based surveillance, wired systems, and others are either *ineffective, inflexible, expensive or impractical* [30], [31], [32]. According to the spillage data from the DPR, using a hybrid system such as the supervisory control and data acquisition (SCADA) systems, detection time can take as long as one month or more. Such long detection times result in greater environmental harm and more financial loss to the operators.



Source: Deloitte analysis.

Deloitte Insights | deloitte.com/insights

Figure 1.6: Digital maturity of the midstream sector [4]

Over the years, technology-driven solutions have addressed many difficulties in various industries and even in our daily lives. Thanks to research, many aspects of technology have continued to grow, ranging from communication protocols to the miniaturisation of devices. One significant growth in recent years is the advent of the Internet of Things (IoT). IoT is the connection or interconnection of machines, objects or devices, humans etc., to the internet. This enables essential features of IoT, i.e. the amassment and analysis of data. The vast amount of collected data has transformed how we live our lives, the procedures of conducting businesses, the kind of service provisions and other extensive arrays of opportunity through efficient analysis [33]. Consequently, more applications of IoT-based solutions are being developed.

While the advent of IoT has been revolutionary and economically impactful, some factors circumscribe its usage. For example, the security and privacy of generated

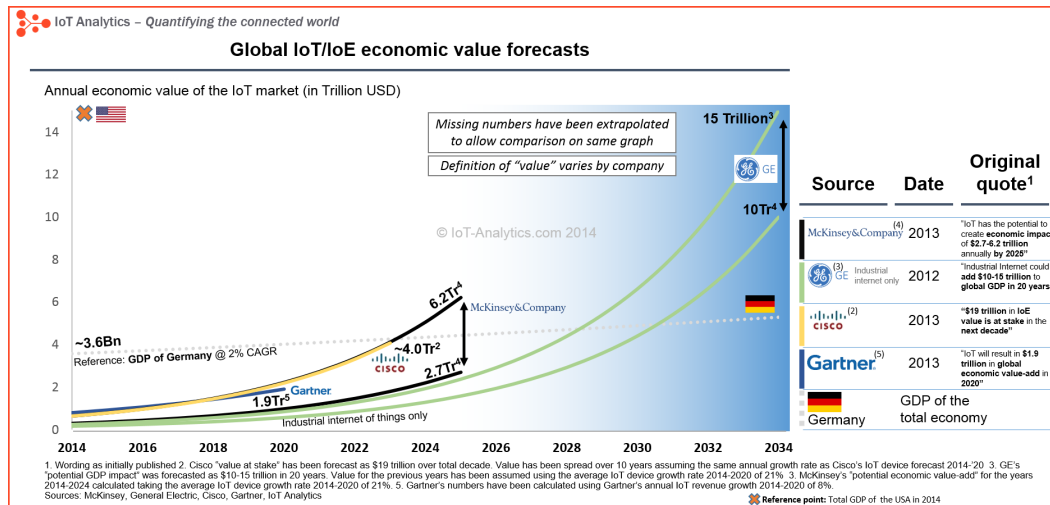


Figure 1.7: Forecasted growth of connected devices

data are vital factors to consider before implementing an IoT-based solution. Other significant factors are the devices' energy consumption, especially their environmental impact, scalability, maintainability, and reliability, among others.

However, the adoption of IoT-based solutions has continued to grow despite the challenges earlier enumerated. Figure 1.7 shows the forecast growth of connected devices in 2034. As shown in Fig. 1.7, this adoption and its projection cut across many industries, such as health, automation, manufacturing and others. Businesses, professionals, and individuals can now make informed decisions using data-driven strategies. Such added values can be seen in several areas, even within a single industry. For example, while Fig. 1.6 shows the current state of digitisation in the midstream sector of OGI, it also shows the different areas, i.e. gathering and processing, pipeline and storage, where changes could be effected, and additional values derived.

Thus, a more recent approach to monitoring pipeline infrastructures is based on Wireless Sensor Networks (WSN) and the Internet of Things (IoT). Although they are proven to be more efficient solutions to pipeline monitoring and fault detection, they are also characterised by some limitations, such as energy consumption, reliability, robustness, and scalability, to mention a few. These challenges have limited most oil and gas companies' implementation of an IoT-based monitoring solution.

However, in our work, we consider an IoT-based solution for addressing the challenges of failures in pipeline transportation of crude oil. Given the challenges of adopting IoT-based solutions, we focus mainly on the following points:

1. How to detect and localise leakages in the pipeline with a high level of accuracy.

2. How to ensure fault tolerance and robustness in the leakage detection monitoring system.
3. How to ensure a scalable yet efficient solution in terms of coverage and connectivity.
4. How to minimise the global energy consumption of the system.

Hence, in the next section, we introduce our contribution to the design of an IoT-based monitoring system for crude oil pipelines.

1.3 Contributions

Our contribution to this work is to provide a fault-tolerant and energy-efficient IoT-based system for monitoring crude oil transportation using pipeline infrastructure. We implement this as an end-to-end solution across the IoT system's different layers (as will be discussed in detail in chapter 3). The first step in our work is to specify the system design and architecture based on which we carry out our implementation and contributions. As explained in the following subsections, our contribution is mainly divided into two aspects (the WSN layer and the Data and Service Layers).

1.3.1 Wireless Sensor Network Layer

The focus of the work in this layer includes event coverage, removal of the Single Points of Failure (SPOF) associated with centralised systems, reduction of false positives and energy consumption compared to existing systems. Hence, we propose a fluid propagation-based node placement strategy to allow distributed leakage detection and localisation. We also implement a detection and localisation algorithm using a hybrid of existing techniques. To test our algorithm, we use NS3 to simulate crude oil propagations. Performance metrics include detection and localisation accuracy, energy consumption, and the number of nodes detecting and localising leakages.

1.3.2 Data and Service Layers

In the Data and Service Layer, we build on the first contribution by extending the detection and localisation of leakages to the fog layer. The main focus of this work is to determine fault-tolerant and energy-efficient data and service management for pipeline transportation of crude oil. To achieve this, we utilise the historical failure of incidents of the Nigerian National Petroleum Corporation (NNPC) pipeline

network. Our proposed solution is a data-driven Markov Decision Process (MDP)-based model to maintain optimal performance across the NNPC pipeline network with minimised global energy consumption through a regionalised approach. We implement three other solutions for comparative analysis: randomised, globalised and pessimistic. Performance metrics include the reward obtained by each algorithm in terms of fault tolerance, the accuracy of detection, and energy consumption.

In the following section, we present the organisation of this thesis.

1.4 Thesis Organisation

This Manuscript is organised into two parts; the first part consists of the background information, and the second includes the contributions. Each part contains several chapters, as discussed next.

1.4.1 Part One: Background

The first part contains two chapters, i.e. the introductory chapter and the literature review.

The first chapter presents the context of our work. It begins by introducing the OGI as well as the challenges and effects of failure associated with the midstream sector. Then we describe the problem statement of our work related to the pipeline transportation of crude oil. Finally, we briefly discuss our proposed solution to the mentioned problems.

In the second chapter, we discuss various research related to our work. This chapter is also divided into two parts. In the first part, we introduce hydraulics and fluid dynamics in pipeline transportation of crude oil. We then briefly discuss various sensors that are used in monitoring pipelines. Also, we delve into the classical methods of pipeline monitoring. In the second part, we examine IoT and WSN-based monitoring systems. In detail, we review works related to strategic wireless sensor placements, their challenges and other aspects of pipeline monitoring. We also look at works on efficiency and data management through information sharing, storage, processing and analytics.

1.4.2 Part Two: Contributions

The second part consists of all our contributions as described next.

In chapter three, we take up aspects of the designs and specifications of our system. We first introduce generic and application-based architectural design and the communication aspects of IoT-based systems. Then we explain the metrics of evaluation on which the system will be designed. These include robustness, reliability, coverage and data management. Finally, we introduce the system design according to the layers discussed in 1.3 and the evaluation metrics.

Chapter four presents our first main contribution. The aspect of this work relates to the WSN layer and presents our detection and localisation algorithm. This algorithm has two versions depending on the degree of connectivity to the neighbouring nodes. We implement this contribution on a single horizontal long-haul pipeline and analyse both versions using NS3- a network simulator. Finally, we discuss and further examine the simulation results and conduct a comparative analysis with other classical methods.

In chapter five, we present our second contribution which extends the work from the previous chapter from the WSN layer to the fog layer, i.e. from a single pipeline segment to a pipeline network. We further introduce an MDP based on the historical failure events of the NNPC pipeline network for efficient data and service management at the fog layer. Additionally, we discuss the implementation and the solution of the MDP, including its impact on the energy consumption of the system.

Chapter six serves as the conclusion chapter. In this chapter, we summarise our work and contributions to this thesis. We also discuss our significant achievements in the course of this research and the limitations of our contributions. Finally, we explore various areas for possible future work.

Chapter 2

State-of-the-Art

Infrastructural monitoring has evolved over the years, moving along with technological advancement. The advent of IoT has further enabled monitoring from the austere environment, such as homes, to complex entities like industrial processes. However, the efficiency of a monitoring system relies heavily on a thorough understanding of the fundamental principles of the entity to be monitored. In addition, it is also essential to study the evolution of other techniques used in the monitoring, such as our use case.

Hence, we present the background on leakage detection and monitoring systems (LDMS) and the current approach to DAL of leakages in two sections- *Classical Pipeline monitoring* and *IoT and WSN-based pipeline monitoring*. In the first section, we focused on the hydraulic background and existing classical work on pipeline monitoring in the background of fluid mechanics in pipeline transportation, the different statistical leakage detection methods and the introduction to the types of sensors used in monitoring pipelines. In the second section, we discuss recent pipeline monitoring techniques based on IoT and WSN, which is in tandem with our contribution. The IoT has several layers, and each layer has different requirements and challenges. Thus the discussions in this section are based on the different layers of IoT, including sensor placement in the WSN layer, data and service placement and their processing.

2.1 Classical monitoring of pipeline transportation of crude oil

Effective monitoring of crude oil transportation in a pipeline can reduce and, in some cases, prevent the negative impact of LAS and other failures attributed to pipeline transportation of crude oil in the midstream sector. This section discusses different types of sensors used in LAS detection and localisation (DAL) in subsection 2.1.1. In subsection 2.1.2, we discuss various fluid mechanics and pipeline transportation phenomena. Then in subsection 2.1.3, we briefly mention some of the classical techniques in DAL of failures in pipelines. Finally, in subsection 2.1.4, we summarise failure detections by their merits and demerits.

2.1.1 Sensors for pipeline monitoring

A sensor is a device that takes an input from a physical quantity and returns an output [34]. Different types of sensors exist and can be used for sensing various physical quantities such as wind, humidity, temperature, and others. In pipelines, leakages can be detected based on the internal changes in the fluid's properties and the external changes as a consequence of leakages and the effect on the immediate surroundings. One or several sensors are required to detect these changes in both cases. In the following subsections, we broadly categorise some of the sensors used to detect and localise leakages and other pipeline failures.

Flow meters

Flow meters measure properties through electrical, mechanical and physical means. We focus on the ultrasonic flow meters typically used in monitoring pipelines. The ultrasonic flow meter consists of a pair of sensors and a transducer, both acting as transceivers to measure the volumetric flow of any fluid. Thus, their internal and external parameters are used to detect leakages in pipelines. Externally, the ultrasonic flow meter depends on the flow clamp-on to the pipeline; internally, it depends on the velocity of the fluid by transmitting ultrasonic waves generated and transmitted from the transducer to the meter (receiver) as well as the temperature of the fluid [35].

The ultrasonic flow meter is either a *Doppler flow meter* or *transit time flow meter*, and both can be used for measuring homogeneous fluid. The transit time flow meter uses the difference in time between sound wave transmission and reception. The Doppler flow meter, on the other, uses the Doppler effect, i.e. the reflection of the sound wave from the transmitter to the receiver. In both cases, measurements are made at several locations and sent to a central system for the algorithmic calculation of the mass volume balance.

While they have several advantages, including robustness, and low maintenance, they also present challenges, such as difficulty detecting small leaks and are expensive to implement [36].

Wired sensors

Wired sensors such as fibre optic cables, conductive or sensing cables and hydrocarbon tubes are used in monitoring pipelines by observing changes such as resistance and impedance in the physical properties of the cable [35]. This is usually achieved by laying them along the length of the pipeline in very close proximity, i.e. in a 10 cm to 15 cm distance.

Fibre optic cables are classified as external pipeline monitoring systems used for distributed chemical, temperature, and acoustic sensing [37]. They use light signals to detect leakages and respond to thermal environmental changes or localised remote vibrations. In addition, they are excellent for underwater pipelines due to their resistance to electromagnetic interference and non-conductibility [38]. Despite this ability for multi-issues detection and low false alarms, fibre optics are not commonly used due to their inability to adapt to pre-existing infrastructure and their exorbitant prices [39]. Additionally, the installation length is limited to a finite distance [35].

Conductive cables consist of a pair of insulated conductors that detect the presence of hydrocarbons by monitoring electrical changes in the cable. These changes occur through the physical contact of hydrocarbon to the cable destroying the insulation between the conductors. This results in changes in current in the cable and thus provides accurate leak location using Ohm's law for measuring the cable's Resistance, Voltage and Current. They can also react to third-party commodities, i.e. non-hydrocarbon fluid, thereby making them susceptible to high false alarms. Although their response time to detection can vary between several minutes to hours, once a cable detects leakage and generates an alarm, it must be replaced, thereby increasing operational expenses. Just like the fibre optic cables, they are also limited to a finite distance, usually a maximum of 400 km [35].

Hydrocarbon sensing tubes detect leakages by sensing the presence of hydrocarbon vapour in the tube that enters it when leakages occur. They can be used for single-phase and multi-phase oil or gas pipelines. The hydrocarbon released during a leak is transported to the tube inlet to the outlet and detected. The hydrocarbon sensing tubes can detect small-sized leakages because of their sensitivity to targeted vapours, and detection time ranges from several minutes or hours to days [35, 37]. However, this sensitivity sometimes leads to false alarms. Just like the fibre optic and sensing cables, they are limited to a finite distance. In this case, a maximum of 50 km.

Visual/Aerial sensors

Remote sensing is achieved using vehicles, satellites, helicopters, and UAVs. to monitor pipelines. They are equipped with specialised sensors like Infrared Cameras, RGB Cameras, and Ground Penetrating Radars (GPR). The visual sensors observe signs of changes or third-party interference in the monitored areas. In the case where leakages are prevalent, community-based and security-based surveillance are employed.

Vehicles are used by security personnel or employed observers like those living within the proximity of pipeline locations to monitor the pipelines. Although ob-

servations like community-based surveillance can sometimes be conducted by foot, it is usually more practical with a vehicle for long-distance pipelines. The vehicles are sometimes equipped with Infrared or RGB cameras for more effective DAL of leakages and spills (LAS).

Satellites are remote sensors equipped with RGB or infrared cameras used to monitor LAS by taking radar images of the monitored area. They can also be recorded and analysed in high temporal frequency. They are efficient for pipelines because they can be used in hard-to-reach areas such as the sea and buried pipelines. Additionally, they are not influenced by the restriction of flight zones as seen in helicopters, thus making them an efficient pipeline monitoring, especially as it relates to their coverage area. Other advantages of satellite-based monitoring include additional monitoring of third-party interference.

Helicopters are used in monitoring LAS in pipelines by flying over the monitored area, usually with an onboard observer(s). LAS is detected or localised with the visual confirmation of the onboard observer. Nowadays, Helicopters are also equipped with various types of cameras, as in satellite-based and vehicle-based monitoring.

UAVs are remotely piloted aircraft integrated with computers and visual sensors. For leakage detection and localisation, they are equipped with forward-looking or high-resolution cameras, synthetic aperture radar, multispectral imaging or shortwave infrared. While this is a cheaper alternative to pipeline monitoring using helicopters, it is currently largely experimental. However, its potential includes the ability to better detect and localise leakages due to its low speed and altitude. It also removes the human-prone risk and errors associated with using pilots and observers in helicopters.

Pipeline Inspection Gauges

Pipeline Inspection Gauges (PIGs) are used to monitor the structural integrity of pipelines and detect defects such as corrosion, dents, cracks, and pits [40]. The monitoring or maintenance operation is done through magnetic flux, eddy current, and geometric detection as the PIG travels in the pipeline. PIGs collect and store information on the pipeline using a battery-powered onboard electronic. More sophisticated PIGs also carry out data analysis and correlation.

PIG can be categorised as gyroscopic-based or non-gyroscopic-based PIGs [41]. In the gyroscopic-based PIGs, the orientation and speed of the device can be easily monitored and maintained, which is not the same for the non-gyroscopic version. However, their deployment is labour-intensive, i.e. using several positioning systems such as visual sensors and odometers. Other challenges include the presence of in-

frastructures, single entrance pipelines, and difficulty in integrating intelligence in PIGs [41, 42].

Pressure sensors

Pressure sensors for crude oil pipeline monitoring are used to measure the flow and speed of the oil in the pipeline. Generally, most sensors use the piezoelectric effect, i.e. the electric charge created by material in response to stress, to measure the pressure. However, some pressure sensors also use resistive, capacitive, MEMS, or optical effects for measurement. The resistive and capacitive pressure sensors measure the changes in the electrical resistance and capacitance of the material, respectively. The MEMS sensors use either a capacitance change or a piezoelectric effect for measurement. Optical pressure sensors, on the hand, use interferometry to measure the pressure of a material. These sensing methods directly influence the reliability or accuracy of the sensors [43].

Pressure sensors come in three different types known as the *absolute*, *gauge*, and *differential* pressure sensor. The absolute pressure sensors are used to measure pressure against an absolute vacuum of zero PSI. The Gauge pressure sensors are used to measure pressure with respect to atmospheric pressure (14.7 PSI). Measurements above or below this value are categorised as positive or negative, respectively. Differential pressure sensors use reference thresholds for measurement. These are the type of pressure sensors typically used for fluid flows in pipelines.

Supervisory Control And Data Acquisition (SCADA)

Supervisory Control And Data Acquisition (SCADA) are used to monitor pipelines by making use of hybrid methods such as the combination of PPA and mass volume balance method.

The placement of flow meters at the inlet and outlet of the pipeline allows leak detection using the mass volume balance method. Flow measurements are periodically sent to the centralised SCADA system, and detection is determined using the *simple volume balance* or the *modified volume balance*. The simple volume balance is wholly based on the principle of mass conservation, where the inlet mass is expected to equal the outlet mass. If the outlet mass is less than the inlet's, a leak is said to have occurred. The modified method includes other state properties, such as the temperature or pressure to determine the presence of leakage. SCADA systems are susceptible to SPOF due to their centralised nature. In addition, they are inflexible to change, have high response time and expensive to maintain or scale [30].

In the next subsection, we discuss the background of fluid mechanics and some computational fluid dynamics and LDTs.

2.1.2 Fluid Mechanics

Fluid transmission can be categorised into different flows like turbulent and laminar flows, steady and unsteady flows, uniform and non-uniform flows, rotational and irrotational flows, compressible and incompressible flows, single direction or multiple directional flows, viscous and inviscid flows, and internal and external flows. Each flow type presents distinct characteristics in terms of changes in velocity and pressure over space and time.

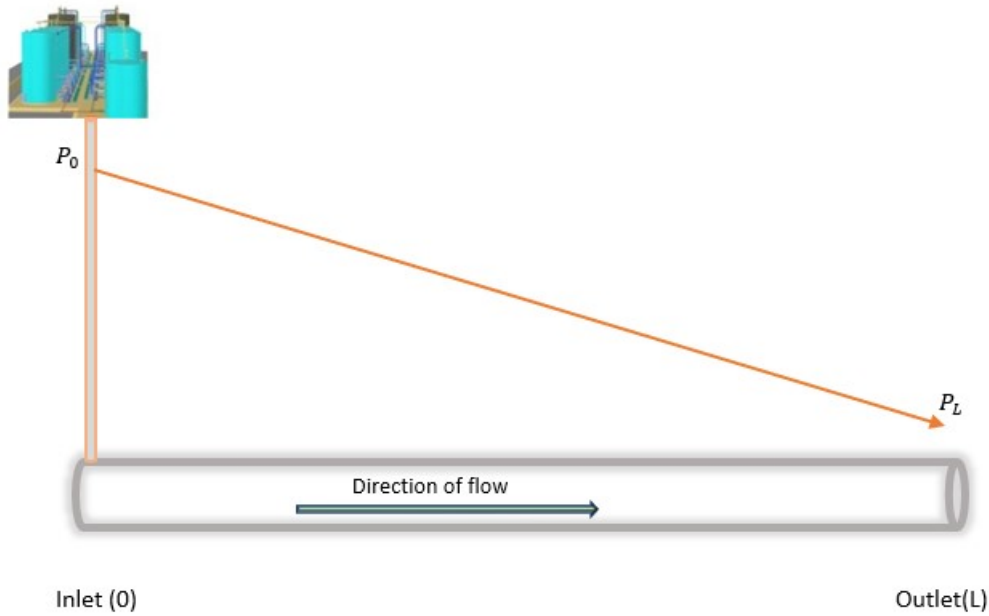


Figure 2.1: Steady fluid transmission in a pipeline

Nevertheless, all fluid transmissions are governed by the principle of fluid dynamics, i.e. the *conservation of energy, conservation of mass, and linear momentum* [44]. Energy conservation follows the second law of thermodynamics, such that the energy rate in the fluid is the sum of the rate of work done on the fluid from the body (and surface) and the net heat flux into the fluid. The conservation of mass states that, the rate at which the mass of fluid changes in a volume is equal to the flow of the fluid. Finally, linear momentum follows Newton's Second Law of motion which states that

the total force acting on the fluid is the mass of the fluid multiplied by its acceleration.

Thus, fluid properties such as the pressure, flow, density, and temperature, are measurable from one point to another following the *equation of state* [45] i.e. any data-driven equation that can be used to describe the fluid phase behaviour [46]. Pipeline transportation of crude oil follows these principles, i.e. for transmission pipelines in a steady state (presenting no leakage), the pressure decreases with distance due to frictional resistance. This decrease in the pressure results in a slope when represented in the time domain as shown in Fig. 2.1.

Changes in the pressure gradient of crude oil travelling in a pipeline are shown in Fig. 2.1. The inlet and outlet pressures in the absence of leakage are represented by P_0 and P_L , respectively. P_0 decreases with distance along the pipeline, hence, maintaining a relatively constant pressure gradient (PG) at every measurement point.

However, when leakages occur, a negative pressure wave (NPW) - travelling in opposite directions from the point of leak [47, 48]- is generated as shown in Fig 2.2. This changes the state of the flow from a steady state to a transient state, after which a steady state is achieved again.

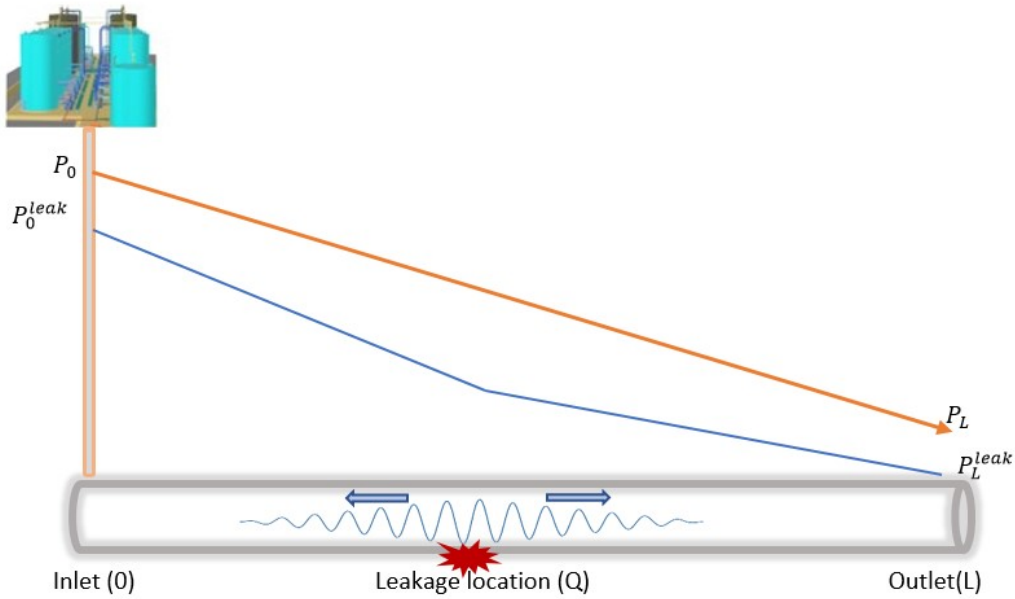


Figure 2.2: Effects of leakage on the pressure gradient and NPW generation

As also shown in Fig. 2.2, leakages result in different pressures at both ends of the pipeline, i.e. P_0^{leak} at the inlet and P_L^{leak} at the outlet of the pipeline compared to P_0

and P_L when there is no leakage. These properties and other changes in temperature and flow rates present useful information in leakage detection.

Computational fluid dynamics (CFD) enables the statistical detection of leakages by advancing the governing principle of fluid flow in either space or time. Nowadays, some of the sensors introduced in 2.1.1 also embeds one or more of CFD method(s), allowing for autonomous detection and localisation of leakage. CFD-based techniques for crude oil pipelines include the pressure point analysis (PPA) technique, negative pressure wave method (NPWM), gradient-based method (GM), mass volume balance technique, and real-time transient modelling [28, 48, 49]. Each technique has advantages and disadvantages (summarised at the end of this section). In the following subsection, we examine these LDTs and focus on the PPA, NPWM, and GM techniques based on their ease of implementation, low computational complexity, robustness, reliability and energy consumption. These properties we have considered are essential in distributed systems, as we will see later in the following chapter.

Pressure Point Analysis (PPA)

The PPA technique is used in determining the dissipation of flow in crude oil pipelines by the periodic measurement of fluid pressure at several points along a pipeline. This is achieved using the well-known Bernoulli equation defined in Eqn. (2.1).

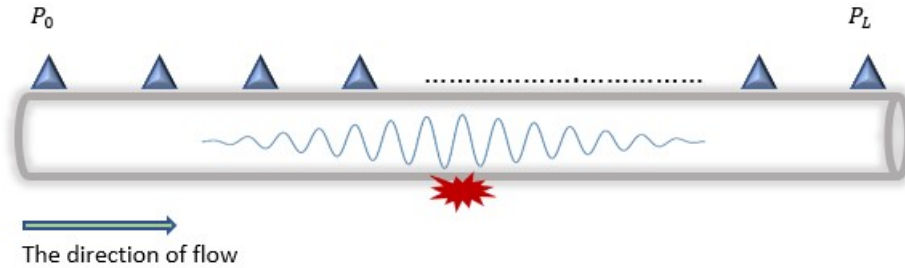


Figure 2.3: Pressure points measurements in a pipeline

$$z_a + \frac{P_a}{\rho g} + \frac{V_a^2}{2g} = z_b + \frac{P_b}{\rho g} + \frac{V_b^2}{2g} + E_{ab} \quad (2.1)$$

where $E_{ab} = \lambda \frac{V^2}{2gd} * L$ represents the energy head loss, λ is the coefficient of friction, V is the velocity of the fluid, d is the pipeline's inner diameter, L is the distance from point a to point b , P_a is the pressure at point a , ρ is the fluid mass density, g is the gravitational force and z_a is the elevation at point a .

Several sensors placed along the pipeline are shown in Fig. 2.3 to make this measurement. The pressure P_i of a sensor i is relative to its location on the pipeline and can be estimated using an inlet pressure such as P_0 . Given P_0 , any other pressure point can be pre-estimated when parameters like the energy head loss E , the fluid velocity and others listed in Eqn. 2.1 are known. Hence, leakage occurrence can be detected by statistically evaluating the measured pressure values against a preset threshold. A resultant mean value less than this threshold indicates a leakage. The PPA technique is easy to implement and is not computationally complex. However, it cannot be used to localise leakages. In practice, it is supplemented with another detection technique for localisation [37].

Gradient-based Method (GM)

The GM technique utilises the changes in pressure gradient when leakage occurs for the DAL of leakages. As shown in Fig. 2.1, the pressure gradient (PG) in pipeline fluid transmission without leakages is approximately the same as the fluid travels along the pipeline. However, two steady states with different PGs are achieved when leakages occur. These steady states are before and after the leak locations, as shown in Fig. 2.4.

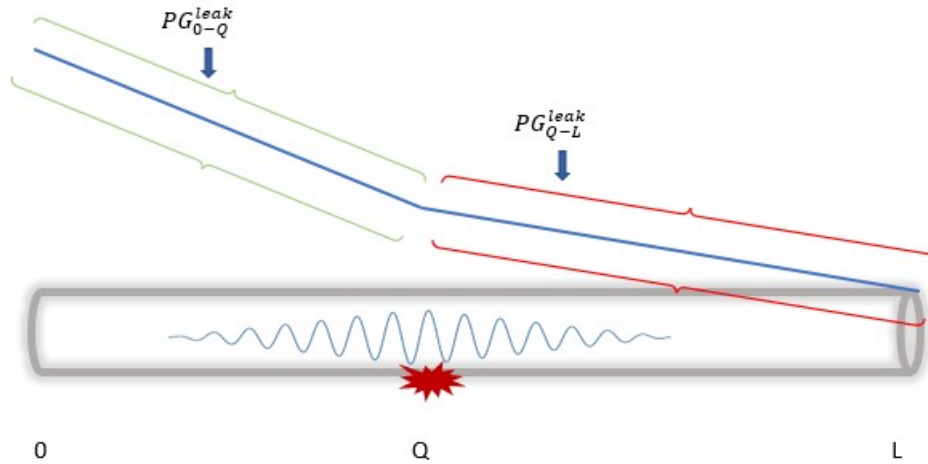


Figure 2.4: The pressure gradient distribution after a leak

$$\overline{Q} = \frac{L \times dPG_{Q-L}^{leak} + (dp_0 - dp_L)}{dPG_{Q-L}^{leak} - dPG_{0-Q}^{leak}} \quad (2.2)$$

where \bar{Q} is the estimated leak location, L is the pipeline length, dp_0 is the average increment in the pipeline's initial cross-section, dp_L =average increment in pipeline final cross-section, dPG_{0-Q}^{leak} is the average increment in the pressure gradient before the leak point, dPG_{Q-L}^{leak} is the average increment in the pressure gradient after the leak point.

For example, if we consider a leakage occurrence at a point Q in the pipeline, and if PG_{a-b} denotes the pressure gradient between two points i.e point a and point b , then PG_{0-Q}^{leak} and PG_{Q-L}^{leak} represent the two gradients before and after the leak point, therefore $PG_{Q-L}^{leak} < PG_{0-L} < PG_{0-Q}^{leak}$. Given this difference in the two steady states, leakage localisation can be achieved with Eqn. 2.2.

The GM technique is an efficient LDT because of its high accuracy, low computational complexity and energy consumption.

Negative Pressure Wave Method (NPWM)

The negative pressure wave method (NPWM) is a commonly used pipeline leakage detection technique [50]. When leakages occur, changes in the fluid pressure shown in Fig. 2.2 at the point of leak generate a NPW. The NPWM takes into account the NPW that is generated when leakage occurs in the fluid pipelines.

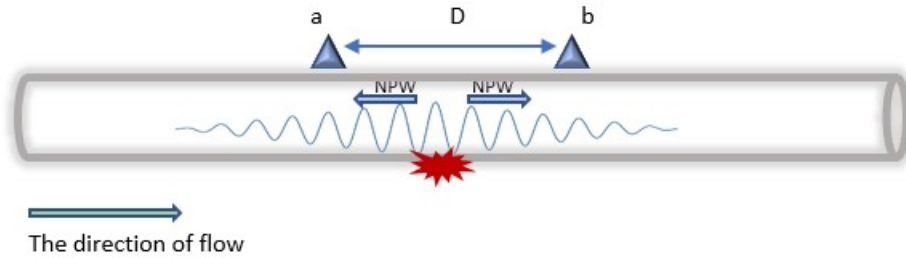


Figure 2.5: Negative pressure wave introduced by a leakage

As shown in Fig. 2.5, the generated wave travels with the speed of sound from the leak point in the opposite direction from the leakage point. This speed (c) can be calculated with Eqn. 2.3. Leakages are characterised by different sizes, from small to large leaks, based on the volume of the liquid lost in a specific period. Hence, depending on the size of the leakage, the amplitude of the NPW attenuates as it travels along the pipeline. Consequently, only sensors within the bounds of the wave can detect the arrival of the NPW front. The corresponding formula for characterising the rate of attenuation of the amplitude of the wave signal is represented in equation

2.4. This formula allows the determination of the maximum distance at which the wavefront is detectable from the point of leakage.

$$c = \frac{1}{\sqrt{\rho(\frac{1}{K} + \frac{d}{Y.w})}} \quad (2.3)$$

where c is the NPW speed, ρ is the fluid density, K is the fluid's modulus of elasticity, d is the pipeline's inside diameter, Y is the Young modulus, and w is the pipeline's wall thickness

$$A_b = A_a * e^{-\alpha D} \quad (2.4)$$

where A_b is the amplitude at sensing point b , A_a is the amplitude at sensing point a , α is the attenuation coefficient, $e = 2.71$ and D is the distance between two sensing points.

$$q = \frac{D - c\delta t}{2} \quad (2.5)$$

where q is the distance from the point of leakage to the nearest downstream node, D is the distance between the sensors, c is the negative wave speed, t is the communication time and δt the difference in the time of arrival of the signal at the upstream and downstream nodes.

The time of arrival of the NPW front can be calculated using Eqn. 2.5 [51]. Each sensor at both sides of the leakage, i.e. the upstream and downstream, receives different times of arrival of the wave depending on their distance from the leak point. Determining the arrival time of this wave is an essential factor in localising leakages. In effect, for NPWM, the size of the leakage and the distance between the sensors are two crucial factors to consider in the DAL of leakages for high accuracy. Generally, NPWM is highly accurate LDT but with high energy consumption due to the sampling rate.

Real-time transient model

RTTM is based on field instrumentation and simulates pipeline monitoring using the hydraulic and thermodynamic properties by measuring the density, flow, pressure, temperature, and the other properties of the fluid. As the name indicates, these measurements are done in realtime representing the state information of the pipeline in all conditions. Important considerations include deciding the boundary conditions i.e., the input data, signal calculation and processing [45].

Mass Volume Balance

Similar to PPA, MVB technique is used to detect leakages when the mass balance at the outlet exceeds a certain threshold defined by Eqn. 2.6 [52]

$$\epsilon \leq \rho_{in} - \rho_{out} - \frac{d}{dt}m_p \quad (2.6)$$

where ϵ is the defined threshold, ρ_{in} is the fluid density at the inlet, ρ_{out} is the fluid density at the outlet, $\frac{d}{dt}m_p$ is the change in pressure and temperature of the pipeline based on the liquid density and the cross-sectional area of the pipe.

The Supervisory and Data Acquisition system (SCADA) typically uses the MVB in combination with other statistical method like PPA for leakage detection and localisation in pipelines.

In the following subsections, we discuss some classical techniques of pipeline monitoring.

2.1.3 Pipeline Failure Detection and Localisation Techniques

The problem of leakage detection and localisation in pipeline networks has been widely studied. Over the years, a multitude of approaches has been proposed as a solution to this problem, with different approaches focusing on addressing specific problems. In this subsection, we work through some of the research conducted in this regard.

Research on robot-based pipeline integrity monitoring is increasingly being carried out. You Na *et al.* [53] proposed using bio-mimetic robots to detect pipeline anomalies. Its advantage is its insect-like crawling ability which enables it to move effectively along complex networks of pipelines. Their work guarantees complete pipeline coverage for pipeline integrity monitoring. Sujatha *et al.* [54] also worked on pipeline monitoring using robots. In this work, a prototype robot aimed at providing continuous and real-time pipeline monitoring in an autonomous manner was built and integrated with a mobile application. The test conducted shows its practicability and promises to be robust and customisable.

Kim *et al.* [55] in their work developed a sensor-based network system for monitoring and maintaining pipeline networks. The system was implemented with the combination of Radio Frequency Identification (RFID) sensors-based localisation technique, MICA-based mobile sensors and topology-aware robots with multi-sensor functions and actuators. It is expected to detect and report anomalies in pipelines before the occurrence of failure and provide recovery or repairs with the help of the robotic agents. The authors supported their ideas by conducting several experiments and

argued that their approach is more cost-effective and scalable than other monitoring systems.

While the use of robots in pipeline monitoring is gaining momentum, it is still at an early stage and is yet to be fully adopted. This could be because of the complexity of their implementations and their performance capability for large-scale or industrial purposes.

Also used in pipeline monitoring are non-intrusive or statistical techniques. Beushausen *et al.* [56] worked on the detection of transient leaks in crude oil pipelines by statistical leak detection method analysing the pressure, modified volume balance and the flow of the crude oil. While some transient leaks were successfully detected in the pipeline, the localisation error was up to 10 km. Other challenges noted by the authors include discrepancies in the flow meter, bandwidth limitation and the effects of the various types of crude oil transported in the pipeline.

Santos *et al.* [57] proposed a non-intrusive leak detection method for liquid pipelines using both Transit-time ultrasonic flow meters and Doppler ultrasonic flow meters. Simulations were carried out to estimate the effects of air bubbles on the efficiency of the proposed system in terms of accuracy in detection and suitability.

In addition, Ostapkowicz, in his work [48], demonstrated two non-intrusive leakage detection techniques based on pressure gradient and negative pressure wave, i.e. GM and NPWM. Experimental results obtained by measuring the pressure and speed of the transmission fluid showed that both methods detect and locate leaks. The GM may be more energy efficient as a result of its low sampling rate. However, lesser accuracy is observed when compared to NPWM.

Karray *et al.* [58] presented a water pipeline monitoring technique based on leak detection predictive Kalman Filter, modified time difference of arrival and System on Chip wireless sensor node. The algorithm incorporates data filtering, preprocessing, and compression aimed at reducing the energy consumption of the nodes.

Mirzaei *et al.* [59] in their work tested the efficiency, precisely the response time in leak detection using Raman Optical Time Domain Reflectometer and Brillouin Optical Time Domain Amplifier. Leak localisation is achieved using the temporal difference of back-scattered laser pulses. Their results showed that the total mechanical response time is several minutes, so they proposed a novel detection based on the results.

Some of the statistical methods, such as GM and NPWM, are highly efficient in terms of accuracy in the detection and localisation of LAS. However, they are implemented in centralised systems such as supervisory control and data acquisition, making them susceptible to Single Point of Failure (SPOF).

Satellites are also used to remotely monitor oil and gas pipelines through image sensing such as radar or RGB images. Kostianoy *et al.* in their works [60–62] shows the efficiency of using satellite monitoring systems for distinguishing anthropogenic and natural effects of pipeline, ports and terminal constructions as well as the ecological impact on the sea within the vicinity of the constructions. Integrated with a sea track web model, the environmental impact of the oil rig was also performed with an analysis of several impact measurements, including their spatial and temporal characteristics. Other works on the usage of satellites for pipeline monitoring include [63, 64]. In the latter, a hyperspectral satellite with the capacity to search for chemicals, gases and other dynamic process on the land. It also enables early detection of pipeline degradation.

While they are especially suitable for difficult terrains like the sea and non-fly zones, they tend to be expensive to implement.

Human-based monitoring systems are also used in the monitoring of pipeline systems. They include using observers in vehicles or helicopters to visually detect LAS in pipelines. Unmanned Aerial Vehicles, such as Drones fitted with multispectral sensors, including infrared and RGB cameras to detect changes unnoticed by human eyes, are also used. Community-based surveillance and Security personnel are also used to detect leakages, especially those due to third-party interference. Although they are well used to a certain degree, studies [30], [31], [32] have shown that these techniques are ineffective, inflexible, expensive or impractical as in some cases failure detection time take as long as five days to multiple weeks. Such a long detection time results in greater loss to the operators with a more severe impact on the environment.

In the following subsection, we summarise various leakage detection systems by their strength and weaknesses.

2.1.4 Summary and Conclusion

Thus far, we have discussed several LDTs, sensors and systems that have been generally classified into external or internal systems, hardware or software-based systems, and human and non-human-based system. In this subsection we enumerate the advantages and disadvantages of the LDTs in Table 2.1 and sensors/systems in Table 2.2.

Detection Method	Category	Advantages	Challenges
Pressure Point Analysis (PPA)	Software, Internal	Leak detection using average pressure measurement, easy to implement	Cannot localise leakage, only suitable for a pipeline in steady conditions
Negative Pressure Wave Method (NPWM)	Software, Internal	Detect small to large leaks in minutes, provides accurate leakage localisation, can determine leak size	Requires high sampling rate, not suitable for long-haul pipelines, event-driven
Gradient-based Method	Software	High accurate leakage detection and localisation, can detect small to big-sized big leaks, low energy consumption	Accuracy tightly dependent on sensors
Mass-volume balance Method	Software	Detect large leaks in minutes	Not suitable for small leaks, prone to false alarms, cannot localise leakages
Real-time transient model	Hardware	Detect large and medium-sized leaks in minutes	Inaccurate leak locations requires multiple online physical measurements, time-consuming, costly

Table 2.1: Summary of pipeline monitoring techniques

Detection Sensors and Systems	Category	Advantages	Challenges
PIG	Hardware, Internal	Can monitor the structural integrity of pipelines	Requires huge labour, difficulty in integrating intelligence
Satellite	Hardware, External	Can be used to effectively monitor subsea pipelines for leakages and spills	Expensive
Community-based Surveillance	Human-based	Provision of security and surveillance in the guarded area	Expensive, ineffective, long detection or localisation time, inability to detect and localise leakages in real-time
Security-based Detection	Human-based	Provision of security and surveillance in the guarded area	Expensive, ineffective, long detection or localisation time, cannot detect in real-time
UAV	Hardware, Human-based	Detect leakages in difficult terrains, Cheaper than Helicopters, lower speed and continuous altitude for improved detection	Long detection and localisation time, limited payload capabilities, susceptible to wind and thermal turbulence interference, limited coverage area and flight time, largely experimental
Ultrasonic Flow Meters	Hardware, External, Internal	Robust, low maintenance	Difficulty in detecting small leaks, accuracy dependent on sensor's sensitivity, does not allow commodity inlet and outlet between sensors
LiDAR	Hardware	Highly effective	Limited coverage area
Infrared Thermography	Hardware	Robust to leakage detection and localisation. Can be used as a hand-held device	Cannot be used for continuous monitoring, usually requires to be mounted on satellite or vehicles for effective monitoring
Helicopters	Hardware, Human-based	Detect leakages in difficult terrains	Inefficient, requires all-round monitoring, covers only small areas, expensive, prone to human errors, high speed reducing the detection capability

Vehicles	Hardware, Human-based	provides security in the areas covered	Expensive, inefficient, prone to human errors
SCADA with Pressure Sensors	Hybrid	Can detect leakages provides valuable information on fluid flow	Expensive, difficult to maintain, inefficient, high false alarms, cannot detect pinhole leakages
Fibre Optic	Wired	Accurate leak detection and localisation detects small to large leaks, can detect leakages in seconds depending on the size	Expensive to deploy and maintain, leakage detection is interrupted when the cable is cut, limited to finite distances
Cables	Wired	Minimal false alarms detect leakage in minutes depending on the size	Expensive to deploy and maintain, not suitable for long haul pipelines
Hydrocarbon sensing tubes	Hardware, Wired	Detects leakages in minutes depending on the size	Cover a maximum of 50km, tube should be fitted in close proximity to a pipeline, expensive, prone to false alarms

Table 2.2: Summary of pipeline monitoring sensors and systems

However, the advent of IoT-based solutions using WSNs has seen a shift in the focus of pipeline failure detection and localisations. A more effective approach, in the next subsection, we look at pipeline monitoring using IoT and WSN-based solutions.

2.2 Industrial IoT and its application to pipeline monitoring

IoT has been used to address different challenges in multiple industries, such as health, supply chain, agriculture, aviation and others. This is made possible with the use of several sensors like pressure, temperature, humidity, or proximity sensors in Wireless Sensor Networks (WSN) and based on several architectural frameworks. Wireless sensors also require strategic placements to cater to specific needs of IoT systems such as energy consumption, scalability, latency, reliability or robustness to failures. Data management in terms of placement, processing and analytics is also a critical factor to consider in IoT-based systems due to IoT devices' limited computational and memory capacity.

Thus, in this section, we present some existing works on the architectures of IoT Systems in subsection 2.2.1. We also discuss wireless sensor placements on pipelines for failure detection and localisation in subsection 2.2.2. In subsection 2.2.3, we examine existing works on WSN-based pipeline monitoring. Then we go into various works and issues on data management in subsection 2.2.4 and associated processing in 2.2.5. Finally, we summarised the challenges in IoT-based pipeline monitoring and the applied approaches in 2.2.6.

2.2.1 Architectural Design of IoT Systems

In a typical communication systems, the system design ranges from the architectural design of the network topology and the communication protocols, to mention a few. Such specification and design phase of any system is the fundamental framework based on which the system is built. It sets the basis for analysis to which we can measure the performance of the system. In IoT-based systems, particularly, the performance of the system is also dependent on such architectural design, the communication protocol, and the data management design, among others. However, unlike the legacy communication systems, there has not been standardisation on IoT's different elements. For example, the number of different layers that form an IoT system and what each layer signifies or constitutes. Many works rely on application-specific designs catering to each application's different needs and peculiarities different from the one shown in Fig. 3.1.

El Hakim, in his work [65], introduced a generic five-layered architecture consisting of sensors or controllers in layer one, gateway devices in layer two, a communication network in layer three, software for data analysis and translation in layer four, and end application service in layer five. Yelmarthi *et al.* also presented a four-layer architecture consisting of sensors, nodes, regional hub and cloud servers from layers one to four, respectively, targeting low-power IoT-based systems. Also, Khan *et al.*, in their work [30], proposed a three-layered centralised architecture consisting of the server control centre, the gateway and the smart object layer for a reliable system. In another recent work [66], Rahimi *et al.*, proposed an eight-layered architecture for 5G and future generation networks. This consist of the physical device, communications, edge computing, data storage, management service, application, collaboration and services, and security layers. These works have presented frameworks for IoT-based systems ranging from three-layers to eight layers. This depicts the high variance in the factors that affect the choices of the system design, Hence, in chapter 3, we take a deeper look at these factors and propose our system design.

In the next subsection, we present various sensor placement strategies in the sensor

layer of the IoT architecture.

2.2.2 Wireless Sensor Placement Strategies

Notably important is the placement of sensors in addition to the choice of sensors adapted for infrastructural monitoring. Sensor placement strategy plays a crucial role in the monitored framework's accuracy, energy consumption, or sensitivity.

Sela *et al.* [67] worked on optimal sensor placement for detection detection on failures on water pipelines. They used a preliminary method of approximate solution of the *minimum set cover* problem based on the *Minimum Test Cover* (MTC) approach. A novel approach based on an augmented greedy MTC-based algorithm was proposed. Conducted test on a water network shows that the algorithm is about three to eight times faster than the other approach. In their other work, Lina *et al.* [68] proposed a robust sensor placement in a pipeline network using robust greedy approximation (RGA) and robust mixed integer optimisation (RMIO). Both propositions served as an enhancement of the nominal GA and MIO using a robustness and redundancy parameter. In most of their simulation conducted on a water pipeline network, MIO and RMIO outperformed the compared versions, i.e. the robust sub-modular function optimisation (RSFO), MIO and GA. Both works are based on the assumption that a single sensor is able to detect failures in multiple pipelines, making them susceptible to SPOFs. [69] is also based on mixed integer programming (MIP) for sensor placements. Results analysis using EPANET, SNL-1, and SNL-2 showed average consensus within the range of 86.5% to 100% with a maximum standard deviation of 1.3%.

On the other hand, Krause *et al.* [70] also worked on robust sensor placement for water networks to avoid intrusions. However, they optimised sensor placement using the minimax criteria instead of MIP. Experimental results include the extension of multi-criterion optimisation, and efficient placement for large networks, i.e. up to 91% of the maximum placement score achievable.

Sarrate *et al.* [71] studied the impact of sensor accuracy based on infrastructural analysis using the isolability index. Application on leakage detection in water networks shows that fault detection is improved and removes the complexity of mesh connectivity in an extensive network.

Boubrima *et al.* [72] in their work for optimal deployment of sensors ensured minimal cost for air pollution monitoring. Two approaches based on integer programming formulation using real air pollution dispersion were proposed, i.e. a basic model and an enhanced model (an extension of the basic model). Both formulations aimed at finding minimum deployment cost solutions through the combination of network coverage, air pollution dispersion and connectivity constraints in a centralised manner. Experimental results showed a considerable reduction in cost.

Guo *et al.* [73] proposed a sensor placement on an oil pipeline to address the sensors' lifetime. They achieved this by taking into account the maximum transmission range of each sensor node. As a result, the distance between the sensors is measured by the length of the pipeline divided by the maximum transmission range of the sensors. The lowest number of sensors are deployed based on their communication range. This approach to sensor placement may significantly increase the impact of event detection if any of the intermediary nodes fail by putting the neighbouring nodes out of reach.

Elnaggar *et al.* [74] worked on sensor placement in a WSN network for oil pipeline monitoring to reduce the impact of energy consumption using ant colony optimisation and genetic algorithm. The simulations conducted on a linear pipeline segment indicate that the ant colony optimisation outperformed the genetic and greedy algorithm in terms of the communication level. However, both approaches show similar behaviour in terms of WSN lifetime optimisation with constraints.

Also based on sensor placement in a linear pipeline is the work of Al Baseer *et al.* [75]. They proposed sensor deployment and grouping based on an adaptive clustering algorithm for intermediate data delivery aimed at reducing energy consumption. Simulation results evaluation shows a significant energy reduction between 300% to over 500% through load sharing mechanism among the cluster heads and up to 62% better than heuristics approaches. Further evaluation using experimental studies shows that their approach conserves energy up to 50% more than the compared scheme.

Li *et al.* [76] proposed a generic sensor placement for sensor network optimisation utilising retransmission and discrete power control for single and double-tier uniformly and non-uniformly distributed WSNs.

In the next subsections, we discuss existing work on pipeline monitoring using WSN.

2.2.3 WSN-based pipeline monitoring

Recent approaches to pipeline monitoring are taking advantage of the advent of WSN and IoT-based solutions. The following works present various approaches to its application for pipeline monitoring and challenges.

Yelmarthi *et al.* [77] proposed a four-layered low-power IoT applications architectural framework from the sensor layer to cloud servers. It comprised easily implementable wired and WSN, with minimal resources for multiple applications. Its applicability in diverse applications and low power consumption were experimentally demonstrated

in damage detection, analysis of posture and physical activities. Khan *et al.* [30] also proposed a three-layered IoT architecture for all the sectors of the OGI. In each layer, they considered reliability and robustness through a hierarchical design. In this structure, interconnection and collaboration enable performance enhancement through reliable communication and intelligent decision-making while allowing predictive maintenance.

Sadeghioon *et al.* [78] proposed a novel algorithm for detecting leakages in underground pipelines through the measurement of relative pressure and temperature obtained from a WSN. In a test conducted, the detection algorithm showed high accuracy in leak detection and sensitivity compared to other threshold-based methods. Also on leak detection is another work of Sadeghioon *et al.* [79]. This research presented a comparative pressure method based on force-sensitive resistors for ultra-low-power wireless sensor networks. Experiments to test this technique were conducted in the laboratory and fields where the leakage was simulated and detected.

Saeed *et al.* [80], on the other hand, worked on a reliable WSN-based system for monitoring oil and gas pipeline that spans over a long distance (REMONG). REMONG specifically focused on how data is sensed and communicated over a sizeable geographical area aimed at reducing energy consumption. A preliminary test for the energy consumed in the communication test showed promising results.

Yunana *et al.* [81] presented a comparative analysis of techniques for monitoring pipeline vandalism. Several monitoring techniques were compared, such as satellite, visual, UAV, and WSN. WSN was more suitable for its low power consumption and cost-effectiveness compared to other techniques. On the other hand, Azubogu *et al.* [82] proposed a WSN-based pipeline monitoring technique. In their work, they discussed several existing monitoring techniques and compared them to WSNs in terms of their architectural design, energy consumption, or maintainability.

Henry *et al.* [83] in their work enumerated the advantages of using WSN for monitoring pipelines in Nigeria, especially vandalism and oil spillage. They focused on different aspects of WSN, including features, challenges and how other applications have utilised WSN. In the work of Obodooze *et al.* [84], they focused on the security challenges related to the use of WSN for pipeline monitoring in the Niger Delta region of Nigeria. Owing to the prevalent vandalism of pipelines, including monitoring equipment in that region, they proposed several measures, i.e. integration of WSN with Wi-Fi network, installation of CCTV, or smart actuators, to circumvent this problem.

Common to all IoT-based applications is also the choice of communication pro-

protocol. Several protocols, such as NB-IoT, BLE, LPWAN etc., exist to cater to the constrained nature of the sensors used in such applications. Thus, we examine some choices in the communication protocols in the following works in terms of their inefficiencies.

Jamali-Rad *et al.* [85], examined the usability of Low Power Wide Area Networks (LPWAN) IoT-based system for seismic quality control. Test conducted on LoRaWAN includes packet loss observation over a small-scaled and single-link network. The performance in the presence of mobility and interference was also tested. Results show that LoraWAN performs well in the presence of interference from LTE, GSM and UTME at the gateway and against Doppler effects. Similarly, Rudes *et al.* [86] conducted a small-scale test on the reliability of LoRaWAN for IoT applications. The test is done with varying parameters such as distance, packet size, and terrain for wildlife detection and precision agriculture. Although there was a lack of optical visibility between the central station and some nodes, obtained results were promising. Ratasiche *et al.* [87] worked on how to ensure a resilient and secure IoT for cyber-physical systems (CPS). In their work, they enumerated various state-of-the-art IoT failure detection and recovery techniques for CPS. Using the presented guidelines, they demonstrated their technique on an application for the mobile autonomous system by using a self-healing approach through structural adaptation (SHSA), which was integrated into the fog node. They capitalised on the redundancy properties and knowledge base of SHSA to monitor, diagnose and recover from anomalies in the communication network.

Sensor devices generate a vast amount of heterogeneous and complex data hence the term Big data. In the IoT value chain, data management and their applications hold between 30% to 60% [88] value. These values lie in the insights obtained through the analysis and optimisation of the generated data. Given the nature of these data, the several constraints of IoT nodes (i.e. memory capacity, computational ability, the limited bandwidth of IoT communication protocols etc., efficient data management in terms of placement, storage, processing and analytics is imperative. In the following subsection, we review works addressing these challenges.

2.2.4 Data and Service Placement

The increasing volume and structure of data, or applications of IoT-based systems, has necessitated alternative storage, accessing, and placement methods as opposed to classical database management systems (DBMS). Traditionally, data and databases are managed with relational DBMS like Oracle, and MySQL, characterised by static schema with records stored in tables [89]. Research [90–92] shows that these DBMSs are not scalable and are not suitable for the management of IoT data due to the

real-time nature of some IoT applications and how the data are collected, i.e. frequency of sampling and heterogeneity. Alternative choices are NoSQL and the so-called NewSQL. NoSQL is scalable, accessible and supports unstructured data, making it more suitable for cloud-based IoT applications or cloud computing in general. Cloud computing provides flexibility and alternative storage maintenance to relational databases, often as platform as a service (PaaS), software as a service (SaaS) and infrastructure as a service (IaaS).

Although cloud computing makes computer resources readily available, the delay for real-time applications, as perceived by the users, continues to be a problem. Content Distribution Networks (CDN) or fog computation are some of the approaches currently adopted to address this QoS issue. Fog computation extends the cloud computing paradigm to the edge of the network, i.e. closer to the users. This enables improved accessibility to the end users' resources, i.e. storage, computation and others. It can also be used to address some of the limitations of centralised cloud computing [93], i.e. by significantly increasing the scalability of the network through the reduction of latency and computational overhead at the cloud server, enhancement of real-time operations, to mention a few.

Despite the many advantages of fog-enabled architectures, misplacing data in the fog nodes can negatively impact the system's overall performance.

Hence, Naas *et al.* [94] proposed a runtime data placement algorithm based on the criteria, i.e. the nature of the data, the node behaviour and location. The results show that overall latency was reduced by 86% compared to cloud solutions and 60% compared to simple fog solutions. Aral *et al.* [95], on the other hand, worked on a replica placement algorithm by taking into account constraints such as the size, location and priced storage. The objective of this strategy is to reduce the latency in fog networks. Results of their work yielded a 14% and 26% -based on tradeoff- reduction in latency compared to replicas' absence.

Additionally, Shao *et al.* [96] also worked on the placement of data replicas for IoT workflows in both fog and cloud environments. In their study, they utilised an intelligent swam optimisation algorithm following several criteria, such as user groups, data reliability, and workflows. Results show improvement in comparison to other works.

Whereas Wang *et al.* [97] made use of multiple channels optimal data scheduling policy in a four-layer fog computing architecture comprising the device, data scheduler, Jstorm, and cloud layers. Big data is split into several blocks and sent to the different Jstorm clusters for processing. The Jstorm layer integrates geographically distributed fog nodes into several clusters. Simulation shows a 15% gain over other data scheduling policies.

In addition to data placement, service placement is just as crucial in a fog en-

vironment, i.e. wrongful service placements also lead to an increase in latency [98]. Hence, Velasquez *et al.* [98] defines an IoT service placement architecture that fuses cloud and fog computing based on the system's condition and latency. In the proposed architecture, services were placed according to the user's location, the server's location and the state of the network while taking advantage of the fog environment. The placement algorithm is generic for all types of scenarios using three modules; service repository, information collection, and service orchestrator.

In the next subsection, we discuss various data processing and management approaches using intelligent algorithms.

2.2.5 Processing, learning and analytics

On top of a cloud and fog-based architectures are several other strategies for improving the overall efficiency such as, reliability, fault tolerance, robustness, responsiveness, scalability, or intelligence of the system. These include data dissemination strategies, intelligent algorithms, middleware and others. For example, the use of microservice architecture can considerably increase the scalability of an IoT network [99]. Unlike its counterpart, it allows distributed services while limiting the functional dependency of the nodes.

Middleware can also be used to provide better functionality in terms of reliability, scalability, availability security or communication [100]. In this work, the importance of service-oriented middleware in the integration and utilisation of fog and cloud computing in a smart city context using SmartCityWare as a service-oriented model was highlighted. Used as a level of abstraction of services and components in smart cities applications, it enhances the flexibility and integration of different services. Experimental results showed reduced latency in fog nodes lookup compared to cloud servers lookup. Ozeer *et al.* [101] also demonstrated the importance of fault tolerance in fog nodes for the provision of reliable services in IoT applications by considering the dynamic, heterogeneous and cyber-physical interaction properties of the fog environment. They proposed a fault management protocol for stateful IoT applications consisting of state saving, monitoring/failure detection, failure notification/reconfiguration and decision/recovery processes. They evaluated their design in a smart home application by introducing a set of simulated events into the application, provoking failure of some of the components and verifying that recovery is as expected with respect to the physical world.

Additionally, a system can be made end-to-end robust, resilient, adaptive and dynamic also by using artificial intelligence/machine learning to extract the unique features of the collected data. Giordano *et al.* [102] used smart agents for the imple-

mentation of self-healing and recovery in autonomous systems through redundancy. In their work, failed or disrupted nodes are replaced with previously available but redundant nodes. According to work done by Nawaratne *et al.*, [33], data interoperability can be achieved through an intelligent algorithm that can self-evolve and adapt, learn incrementally with temporal changes and have unsupervised self-learning capability. Also, Tang *et al.* [103] extended the data analytic capability of cloud computing in the context of smart cities using smart pipeline monitoring prototypes.

Besides, the advent of federated learning supports machine learning at the fog/edge nodes [104], [105]. This also encourages better performance by bringing intelligence closer to the originating data point and limiting the amount of data that needs to be transmitted to and processed in the cloud. As a result, failures that may have arisen due to network failure or overload are reduced. Rashid *et al.* [26] used machine learning for the leakage detection and classification in pipelines. For their experimental set-up, sensor nodes were equally spaced along the pipeline based on their communication range. In this work, several machine learning algorithms, such as K-nearest neighbour (KNN), support vector machine (SVM), and Gaussian mixture model (GMM), were compared. They outperformed the rest of the algorithms regarding specificity, sensitivity, and accuracy of leak size estimation. Likewise, Roy [106] worked on an artificial neural network consisting of input, hidden and output layers. It is used as an optimisation tool for accurately detecting leakages in pipelines.

In addition, employing necessary telemetry for performance monitoring can significantly improve reliability through context management, i.e. a device is dedicated to monitoring a predefined set of metrics like link quality or battery level before tasks are allocated [93]. The proposed services can be implemented through virtualisation to further enhance system reliability. Dar *et al.* [107] proposed a cloud-based IoT service that ensures reliability through a framework used in defining systems requirements and dependability demands. A virtualisation service is called during runtime that deploys one of the predefined redundancy patterns applied. Experimental work showed that application of this method allows over 98% availability with nearly 40% randomised network failures. Downsampling transmitted data also ensures reliability by limiting bottlenecks, thereby minimising packet loss rate. This can be done through efficient data transmission algorithms, e.g. data fusion and data prioritising algorithms as proposed in [108] to minimise packet loss rate due to network overload or limited bandwidth.

Other important considerations include the information flow between the network layers. Most commonly used are the publish-subscribe (pub/sub) messaging paradigms between clients (data producers) and agents (data consumers). Ioana

et al. in their work [109] demonstrated the applicability of the pub/subsystems in several complex scenarios, such as the Open Platform Communication Unified Architecture protocol for industry 4.0. In particular, they showed how a multi-channel User Datagram Protocol (UDP) communication strategy for pub/sub systems enables the transmission of high-volume data like images in a time frame fitted for the industry. Aslam *et al.* [110] worked on adaptive methods to handle unknown subscriptions in a low-latency pub/sub model used for processing multimedia events. Their system achieved between 79% and 84% accuracy. An interesting work [111] also on the pub/sub systems focused on minimising the computing and transmission of cost for content subscribers based on a requested format.

Some works for latency reduction in mobile edge computing and general optimisation are based on game theory. In their work [112], Garg *et al.* evaluated three dynamic placements (greedy approximation, integer programming optimisation and learning-based algorithms) for maximal user equipment availability using minimal infrastructures. Experimental results using a drone swarm application shows that while all approaches met the latency requirement, the learning-based algorithm performed better in terms of minimal variations in solution providing a more stable deployment and thereby guaranteeing a reduction in infrastructural cost. Also on placement optimisation is the work Ting *et al.* [113]. They worked on the optimal provision of edge services such as storage, communication and computational resources. Using a trace-driven simulation, they compared results obtained on the following algorithms: Optimal request scheduling (ORS as the baseline), greedy service placement with optimal request scheduling (GSP ORS) and greedy service placement with greedy request scheduling (GSP GRS). For joint service placement and resource scheduling, both GSP ORS and GSP GRS, including their linear programming relaxation, either achieved optimal or near-optimal solutions.

Improving data management through a game theoretical approach is also being researched. Cai *et al.*, in their work [114], proposed a Reinforcement Learning Heuristic Optimisation (RLHO) framework aimed at the provision of better initial values for the heuristic algorithm. They carried out a comparative analysis between their proposed algorithm and two other algorithms (Simulated Annealing and Proximal Policy Optimisation). Results obtained from experiments on bin the packing problem show that RLHO outperformed pure reinforcement learning. Islam *et al.* [115] and Rezazadeh *et al.* [116] worked on third-party interference on pipeline networks, especially a terrorist attack. The former proposed a Stackelberg competition-based attacker-defender model to find the equilibrium between possible attacks and pipeline security. They proved that in an equilibrium state, the monitoring system achieves the best result by maintaining its strategy, assuming both the defender and attacker act rationally. The later [116] proposed another model of the problem using a two-player non-zero-

sum approach, with the assumption that both players act rationally according to some chosen indices. Two approaches to solving the problem were proposed. The first is a local optimisation approach allowing comprehensive analysis of the effects of countermeasures on attacks. The second approach utilises a global optimisation approach that enables the security personnel (defender) to provide a solution from the attacker's perspective.

Rezazadeh *et al.* [117] also worked on modelling a monitoring system for pipeline security using the Bayesian Stackelberg game. In his work, he proposed a time and distance discretisation-based scheduling policy. This framework enables the hierarchical ranking of security risks allowing the usage of different patrol paths.

So far, we have discussed multiple works towards efficient pipeline monitoring, data management and processing. In the next subsection, we will summarise the different challenges in the adoption of WSN and IoT-based solutions for infrastructural monitoring.

2.2.6 Summary and Conclusion

One factor to consider in utilising IoT-based solutions for infrastructural monitoring is the multiple limitations of any IoT-based system. In the preceding subsections, we elaborated on many of these issues and works done to circumvent them. Therefore, we summarise these challenges and possible solutions in this subsection.

Challenges	Approach applied
Reliability	Hierarchical and distributed architecture, self-healing through structural adaptation, multi-objective QoS routing, retransmission, intermediate node retry
Energy consumption	Protocol selection, device deployment strategy, data sensing strategies, optimised QOS multipath routing, data caching and proxying, aggregation, anycasting
Scalability	Fog computation and hierarchical architecture, software-defined networking
Robustness and Fault Tolerance	Middleware services, distributed architecture, interest aggregation, anycasting
Data, topology, and coverage optimisation	Federated learning, particle swarm optimisation, Gaussian distribution, ant lion optimisation, differential evolutionary and genetic algorithm
Latency	Fog-based and hierarchical architectures, data placement, service placement, middleware services, network function virtualisation

Analytics		Federated learning
Sensor	place-	Minimum Test Cover, robust mixed-integer optimisation, mixed
ment		integer programming, isobility index, sensor transmission range,
		colony optimisation, adaptive clustering
Data and service		Runtime data placement, intelligent swam optimisation, net-
placement		work state, user server location

Table 2.3: Summary of common IoT challenges in pipeline monitoring and applied approaches

Part II

CONTRIBUTIONS

Chapter 3

System Design and Specifications

Several options exist in the design and deployment of IoT-based systems for monitoring industrial processes. Hence, in this chapter, we discuss the design choices of our system compared to generic standards and existing models in other application areas. In section 3.1, we discuss some generic and application-based architectural designs for IoT systems. Additionally, we present some commonly used communication protocols in these systems. Section 3.2 shows and discusses the metrics on which some of the design decisions are based. Finally, in section 3.3, we introduce our proposed system architecture and its various components.

3.1 Background

The architectural layout, communication protocol, and other elements of a communication system have several impacts on the system. It can affect performance metrics such as robustness, reliability, and maintainability, to name a few. This section introduces some common architectures and communication protocols for IoT systems and their impacts.

3.1.1 IoT Generic and Application-based Architectures

We present a generic IoT architecture in Fig. 3.1 as proposed and explained by the International Telecommunication Union in their recommendation report [5]. This architecture has four layers consisting of the device, network, service and application supports, and application layers. Each layer is further embedded with management and security capabilities.

At the device layer, two general categorisations exist, i.e. device and gateway capabilities. The device's capabilities include interacting with the communication network directly, indirectly or through ad-hoc networking. The gateway capabilities, in addition to providing indirect communication, can support multiple interfaces and protocol conversion.

The network layer mainly deals with networking connectivity through the provisioning of control functions, mobility management and others. It also deals with the transporting aspects for services, and application data, to mention a few.

At the service support and application layer, two types of categories are present, i.e. the generic and specific support capabilities. While the latter includes supports

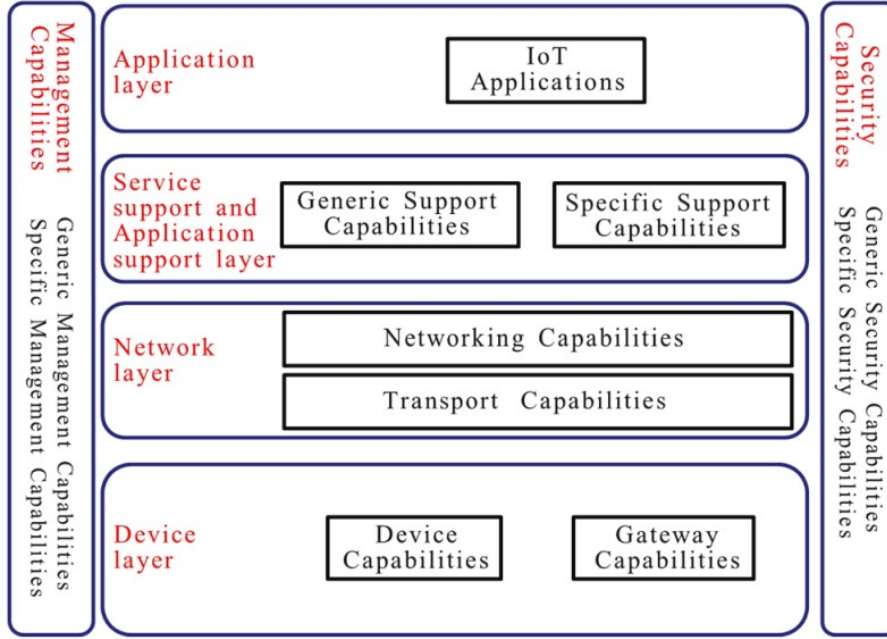


Figure 3.1: Generic IoT Architecture [5]

such as data processing or storage, the former deals with supports unique to different applications. Finally, the application layer contains all applications for the different subject areas where IoT-based solutions are applied.

The management capabilities cut across the four layers dealing with the network topology, the traffic controls and the general management of all devices. The security capabilities, also dealing with all four layers, include authentication, authorisation, and data confidentiality across these layers.

The generic architecture presented in Fig. 3.1 aims to cover all the aspects of an IoT-based system. A typical system includes devices - that could act as sensors, gateways or both- that are used to percept the physical world. In addition, the networking aspects, such as the communication between these devices for data transfer and information interchange, are imperative. As such, several communication protocols are considered in this aspect, typically catering to constrained devices like IoT-based devices or specific application scenario. Various specifications and choices are made based on the application or the industry which the IoT-based system is deployed.

We have also shown in chapter 2-by presenting multiple existing works- that there is no unified architectural design that has been generally adopted for IoT-based monitoring systems. Most architectural frameworks are dependent on various factors,

including the use case and the specific outcome required. However, paramount in the design choices is the communication protocols.

As such, we propose a system design considering factors such as the industry of application, the availability in place of deployment (POD), existing communication infrastructure in the POD and possible integration with such infrastructure. Mainly, the design aspect of our system is aimed to further enhance its performance in several folds, as will be seen later. The design aspect of work is part of our published work in [118]. Thus, in the following subsection, we briefly discuss some communication protocols used in IoT systems. Other factors affecting the design or deployment of IoT-based can be found in chapter 2.

3.1.2 Wireless Communication Protocols for IoT

IoT/WSN sensor nodes are by nature constrained devices limited by their computational ability, memory capacity, energy consumption, security and privacy capabilities, among others. Existing cellular communication networks such as the 3G, 4G, LTE and others do not provide an energy-efficient utilisation of the IoT devices [6]. As shown in Fig. 3.2, other wireless technology and standards currently exist and meet the requirements or needs of IoT-based systems. Mainly differing by their communication range, which could range from zero (proximity) to up to a hundred kilometres and the throughput. Such wireless technologies are typically broadly categorized as short and long-range protocols.

The short-range communication protocols include IEEE 802.15.4 (Zigbee), IEEE 802.15.1 (Bluetooth LE), 6LowPan, Wi-Fi and others. Zigbee and Bluetooth LE shown in Fig. 3.2, are both examples of WPAN with a maximum transmission range of about 100m. The data rate of Zigbee is approximately 250kbps, and it has an application throughput of 20kbps [119]. Zigbee also uses mesh topology for data transmission. 6LowPAN, similar to Zigbee, also utilises a mesh network topology with a throughput of 20kbps and does not make use of battery. Conversely, Bluetooth LE has a point-to-point topology in a master-slave setting where the master is responsible for communication management and the slave execution of commands. In addition, the Bluetooth LE uses a coin cell battery which can last up to a couple of years.

The long-range protocols, i.e. Low Power Wide Area Network protocols (LPWAN), can be further categorised into licensed and unlicensed LPWAN [120]. LoRaWAN and Sigfox are examples of unlicensed LPWAN, while the 3rd Generation Partnership Project(3GPP) NB-IoT is a licensed LPWAN. LoRaWAN operates in different radio bands across the continent, e.g. it uses 868 MHz in Europe. With a low data rate, the throughput is dependent on several factors, such as the size of the

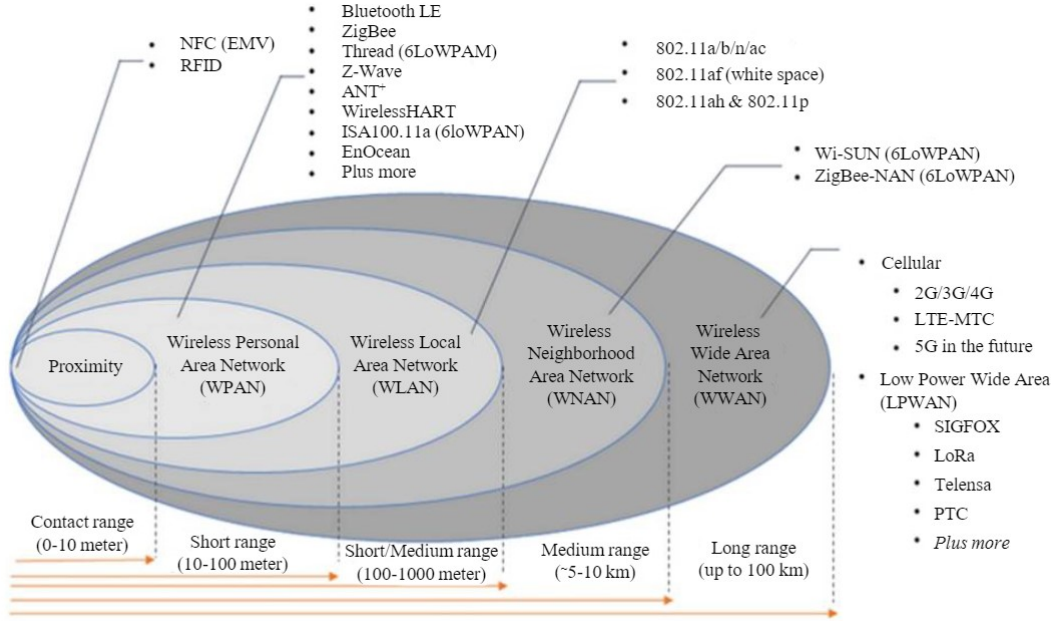


Figure 3.2: Wireless Communication Technologies [6]

bandwidth and the spreading factor. Sigfox, also operating on the 868 MHz band in Europe, uses an Ultra Narrowband signal with low energy consumption. It is limited to a certain number of messages delivered daily, depending on the data rate, up to 100bps. NB-IoT is a 3GPP standard and like the name implies, operates on a licensed spectrum of 200kHz band. It has a connection density of up to one hundred thousand devices and it is energy-efficient.

Recall, that this work focuses on detecting and localising leakages using an efficient IoT-based LDMS for a pipeline network spanning several thousands of kilometers in Nigeria. To design such system, we consider some of the performance-affecting factors of the communication protocols in terms of coverage, availability in the POD as well as the energy consumption to ensure longevity of the design. Thus, Table 3.1 summarises the properties of some commonly used and new communication protocols for IoT-based systems.

In the next section, we discuss the design metrics considered for our system and on which the performance of the system will be evaluated in the following chapters.

Protocol	Range	Data rate	Topology	Availability in PoD
BLE	100m	125 kbps-2 Mbps	Mesh, scatternet, P2P	Yes
ZigBee	10-100m	250Kbps	Mesh, star, tree	Yes
6LowPAN	10-100m	200kbps	Star, mesh	Not yet
SigFox	10-50km	100bps	One-hop star	Not yet
LoRaWAN	5-45km	250bps-50kbps	Star, star of stars	Yes
Narrowband IoT	1-10km	100-600 bps	Star	Not yet
Weightless	10km	0.1 Mbps to 16Mbps	Machine to machine	Not yet
MIoTy	5km -20km	400bps	Star	Not yet
Wi-Fi (802.11n)	50m	72.2-450Mbps	Star, Mesh	Yes
RFID	Proximity	40-640 Kbps	Point to Point	Yes
NFC	Proximity	106 to 424 kbit/s	Point to Point	Yes

Table 3.1: Properties of some IoT Wireless Communication Protocols

3.2 Design Metrics and Considerations

A mission-critical IoT system consists of heterogeneous devices, data and applications whose failure may cause highly impacting consequences on the environment, public services and others. Additionally, it requires high availability, robustness and reliability, to mention a few [121]. Based on this definition, a pipeline infrastructure monitoring system based on IoT can be categorised as a mission-critical IoT system. Thus, we aim to design an end-to-end resilient IoT-based monitoring and detection solution that is robust to failure and efficient in terms of energy consumption and coverage. The system will also address some of the identified challenges in the pipeline infrastructure of the oil and gas midstream sector. These challenges include high false alarm rates associated with traditional LDMS, complex or computationally expen-

sive detection techniques, SPOF related to centralised systems, poor detection and localisation accuracy, and high energy consumption as enumerated in the following works [47, 48, 55, 67, 79].

To design such a resilient IoT-based monitoring system, we consider the challenges of pipeline infrastructure, the requirements of a mission-critical IoT system, and the existing solutions and protocols introduced in 3.1.

Therefore, our objectives for the system design can be summarised as follows:

1. Ensure reliability and robustness.
2. Ensure coverage and connectivity.
3. Ensure highly accurate leakage detection and localisation.
4. Minimise energy consumption.

These objectives also served as the metrics of evaluation for our system. In the following subsections, we explain some of the design criteria and common approaches to their design.

3.2.1 Reliability and robustness

By definition, a resilient system is a system that is dependable in the presence of all types of faults with the ability to evolve or adapt to different situations. It should be *scalable, available, safe, maintainable and reliable*. In addition, a robust system is a system that can continuously function in the presence of stochastic interference [87]. Note that, failures in pipeline infrastructure cannot be predetermined, i.e. they are stochastic in nature. In addition, they often include third-party interferences resulting in the disruption of installed monitoring systems. Thus in our case, the key challenges to address are the availability and reliability of the system. Several techniques have been used in system designs to improve the availability and reliability of any system, such as nodes and channel redundancy. The redundancy technique can significantly improve recovery time and infrastructure options when a failure occurs. Although it incurs an extra cost, it allows continuous operation and delivery of service in the event of failure [87]. Other robustness techniques include employing fail-silent security measures to detect unwanted interference in the system.

In a distributed system, the degree of distributiveness, application-level checkpoints, and the architectural design are also used to provide resilience in system designs [122–125]. A hierarchical and distributed network architecture can also allow for a scalable network without substantially affecting performance metrics like throughput, latency, and energy efficiency, among others.

3.2.2 Coverage and connectivity

Coverage can be considered in two aspects: the physical or event aspect and the communication aspect. The physical aspect, i.e. detecting changes in the monitored quantity, is one of the crucial factors to consider in system design. For example, Boubrima *et al.* [72] modelled air pollution detection based on the Gaussian air pollution dispersion in the environment. With this, he jointly proposed an integer formulation model of the dispersion for coverage and connectivity constraints. Alam *et al.* [126] also worked even detection, focusing on dynamically changing the sensor range. Other method includes density-based and hot-spot-based estimation as in [127]. In our work, the detection of such physical quantity is based on the changes observed when leakages occur. As discussed in more detail in chapter 4, we specifically considered the NPW and PG generated in the event of leakage.

The second coverage aspect concerns network coverage ensuring end-to-end communication among the devices. The issue of network coverage can be examined in terms of the network topology, the communication protocol and the availability of the protocol in the place of deployment (POD).

3.2.3 Accuracy of leakage detection and localisation

Accuracy of leakage detection and localisation (DAL) is one of the determining factors considered by oil and gas operators when considering leakage detection techniques or monitoring systems [48]. Incorrect or insufficient data can lead to high false alarms, as it is with some leakage detection techniques. Thus, in addition to event coverage, we consider the correctness of the information or data collected. This can greatly influence the accuracy of the output result. In addition, accuracy can also be determined with the use of machine learning techniques such as deep neural networks or convolutional networks. In [128], a data-driven neural network model was proposed for accurate text recognition. Wu and Lin [129] also used convolutional neural network for accuracy in waveform inversion.

In our work, the accuracy of DAL is determined through the fusion of multiple detection techniques in more than a single node. A detailed explanation of the method is given in the following chapter 4.

3.2.4 Energy consumption

There has been a rise in awareness of the negative impacts of energy consumption on the environment. Hence, the reduction in the overall energy consumption of a system has become paramount in any system design. An IoT solution's lifespan and

performance depend highly on the nodes' energy efficiency. Several measures have been taken to minimise the energy consumption of IoT devices and systems. These include the use of energy-efficient communication protocols, the architectural design of the system, and the node deployment style, among others. Li *et al.* proposed the minimisation of the energy consumption in their system by implementing an efficient data collection strategy and the system's network design. In some work, energy harvesting schemes are employed to preserve and improve the life cycle of the devices. Management of the nodes and network activities are also employed to improve the energy consumed in system design.

3.3 System design and discussions

In our system design and specification, we put into consideration all discussions in the previous section. We begin by defining the system's architectural design, then determine the communication techniques between nodes and layers and finally, the data management aspect of the system.

3.3.1 Architectural design

As discussed in section 3.1, there is no consensual architecture for IoT systems. Recall that in our design, we want a reliable architecture that is efficient in terms of latency, energy consumption and reliability. Thus, we propose a three-layered hierarchical and distributed architecture. The hierarchical nature of the design can ensure a reduction in latency for the real-time detection and localisation of leakages. The distributed nature correspondingly ensures robustness to communication failures, SPOF, node failures and other types of failure that may result from third-party interference or natural occurrences on the pipelines.

Figure 3.3 shows the three-layered architectural design for our system, *the sensor, fog and cloud layers*.

The sensor layer consists of several sensors placed on the pipeline to collect information and data about events on the pipeline. To realise that, we make use of multi-spectral sensors, i.e. sensors with the ability to collect diverse information. Information collected in our case includes the negative pressure wave (NPW), the speed of the NPW, and the pressure on the pipeline. This information aid in detecting and localising leakages.

At the fog layer, we have several gateways for collating sensor information. At this level, more complex and computationally tasks on the collected data are achieved. For example, data management such as prioritisation, classification, placement and replication, automated learning are also done in addition to the detection and localisation

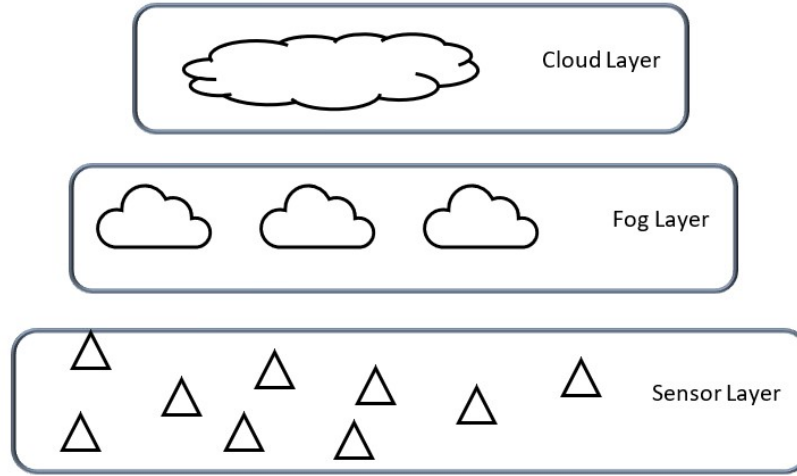


Figure 3.3: Network Architecture

of leakages.

The cloud layer is used for carrying out long-term services such as the storage of historical data, predictive analysis, alarm notification and others. In the next subsection, we discuss the communication aspect of the proposed architecture.

3.3.2 Communication protocols

In section 3.1, we introduced several communication protocols suitable for IoT systems. Each protocol presents different properties regarding communication range, latency, and energy consumption, as listed in Table 3.1. Particularly interesting for our use case is the LoRaWAN network communication protocol. It is a long-range protocol that is power efficient, cost-effective, easy to maintain and configure and scalable. As demonstrated in [85], LoRaWAN is suitable for long-distance communication (more than 30 kilometres), performing excellently well in an out-of-line of sight with rugged terrain.

In terms of scalability, a LoRa base station has about 20km coverage range, and each cell has the capacity to connect up to 50 thousand end devices which allow for a highly scalable network compared to NB-IoT as stated in the comparative studies of the LPWAN technologies [120]. In addition, when considering deployment cost, a LoRa base station costs four times and fifteen times less than its Sigfox and NB-IoT counterparts, respectively [120].

The end devices communicate with each in the sensor layer. The gateways also

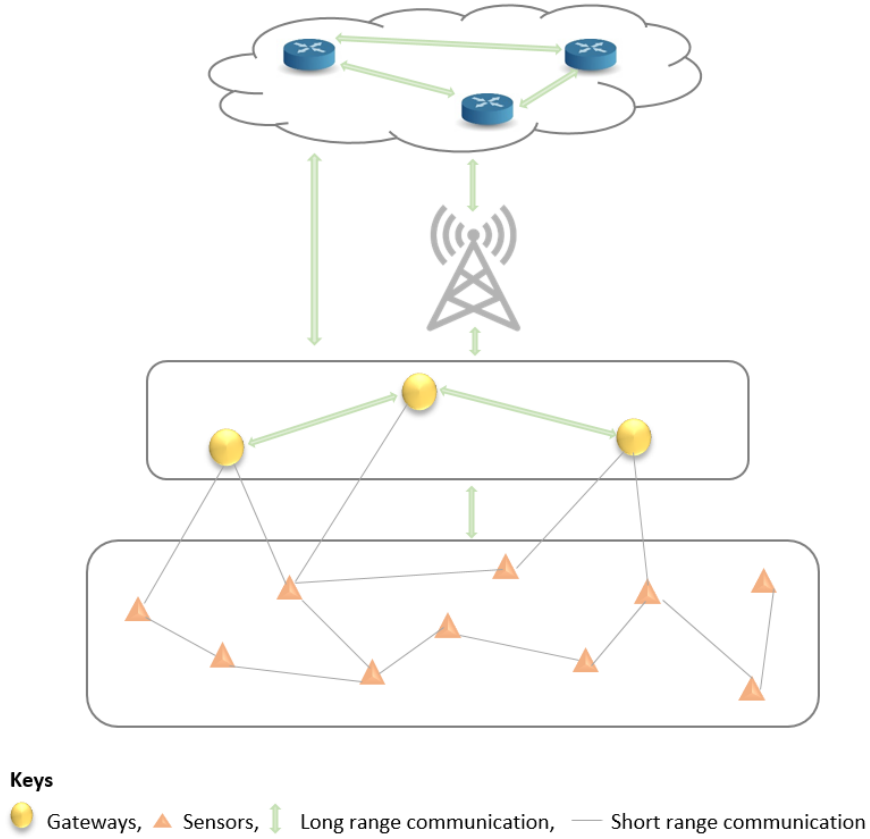


Figure 3.4: Network Architecture with Communication

exchange information, and the same goes for the cloud data centres. Finally, there is interlayer communication, i.e. from the sensors to the gateway, from the gateway to the cloud and vice versa. Typical communication in our system is shown in Fig. 3.4. Thus, we use LoRa for the long-range protocol between end devices, gateway devices and end device to gateway device. As a backhaul for long transmission, we employ cellular networks such as 3G, 4G and LTE (or other available choices). The choice of the short, and long-range protocol and backhaul depends on their availability in the POD as well. In our case, all these technologies are currently available in Nigeria from various operators such as MTN, Globacom, Airtel and also, a communication satellite for those areas that be hard to reach.

The first aspect of coverage, i.e. network coverage, was addressed in the previous subsection. We also mentioned problems related to existing detection techniques and monitoring systems, such as high false alarm rates and poor accuracy in detecting and localising leakage. In the following subsection, we discuss the second part of coverage

in our work.

3.3.3 Event coverage

In this subsection, we briefly introduce our approach to leakage event coverage, taking into account the reduction in false alarm rates, the accuracy of detection and robustness to failure.

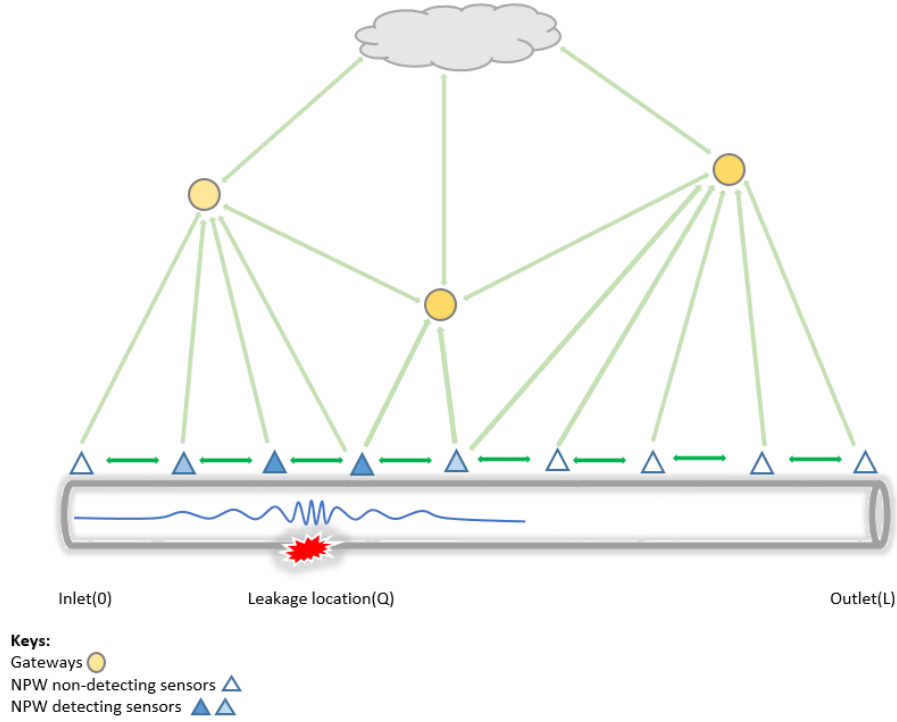


Figure 3.5: Event coverage on the proposed architecture

As shown in Fig. 3.5, node placement on the pipeline is based on the negative pressure wave generated when leakage occurs (more details in chapters 1 and 4). To ensure maximum coverage in both event detection and its transmission between sensors or between sensors and gateways, we implement the following steps:

1. Implement redundancy to allow autonomous recovery of failed or disrupted nodes with a functionally equivalent one. Network route redundancy can also improve robustness in the event of communication or node failures. The inter-node connections are extended to nodes in the same layer in addition to their connection to nodes in the higher or lower layer.

2. Reduce false alarms by ensuring that more than one node can detect the leakage.
3. Reduce the energy consumption due to the architecture's distributed nature, we first determine the leakage region, i.e. where the effects of leakage are measurable and Localise the leakage using only sensors in this region.

In the next subsection, we discuss how the data collected from the leakage events are managed across the network architecture.

3.3.4 Data management

Based on the three-layered network architecture, our data and service layer are structured in three layers across the defined architecture. Fig. 3.6 shows the various data and service placements from the sensor to the cloud level. Data and services are further divided into two sub-layers, using the publish-subscribe paradigm as the model of interaction between the sub-layers. For the management, we consider various aspects, from data creation, communication, aggregation and prioritisation, and storage to data analysis. In our work, the end devices generate various data from the pipeline by periodically carrying out the measurement on the pipeline.

The generated data are made available through publication. Services can then subscribe to the different kinds of data that are of interest to them. Whenever data is published, subscribed service(s) receive the data for their various uses, i.e., the data communication aspect.

The proposed architecture allows data communication or data sharing amongst pre-defined neighbourhood sensors at layer one. Such collaboration among the nodes enables the implementation of services such as the detection and localisation of leakages. Data sharing is also extended to the fog level to enable services such as data preprocessing, prioritisation, placement, aggregation and replication, among others.

While data storage cuts across the three levels (sensor, fog and cloud), each level deals with storage depending on the needs, size and implemented services, with the sensor layer at the bottom of the hierarchy. The fog layer deals with the main part of the work, i.e., the implementation of efficient fault tolerance, which will be shown in chapter 5. The cloud level, on the hand, houses more extensive data for more intensive or complex computations.

More detailed explanations of the system design and specification are given in the following chapters as it relates to the contribution of each chapter.

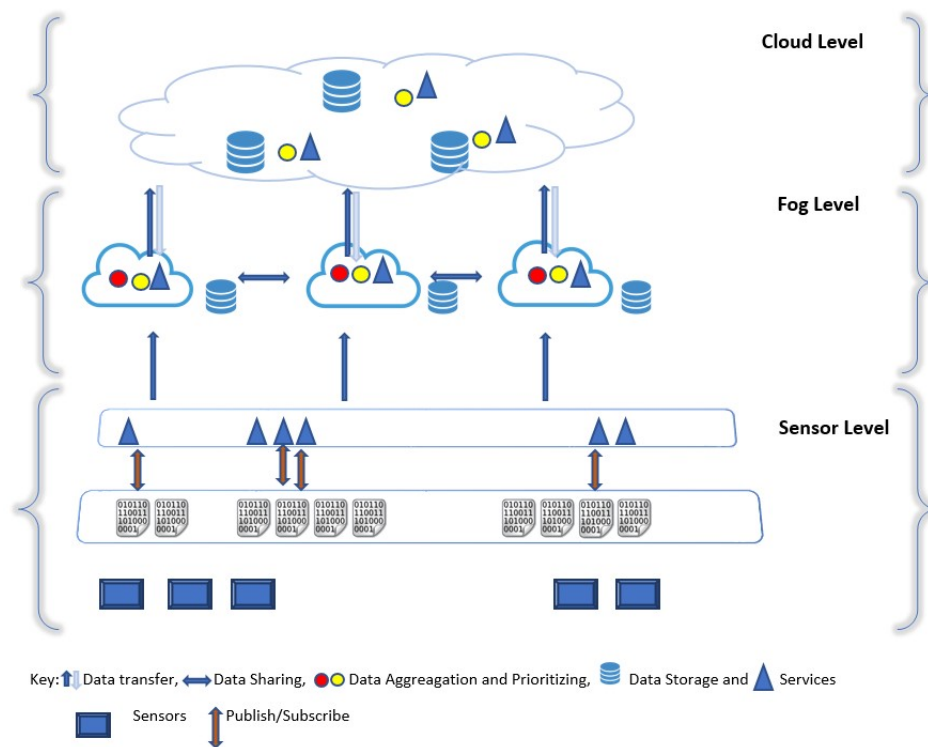


Figure 3.6: Data and Service Placement

Chapter 4

Distributed Detection and Localisation of Leakages in Pipelines

In recent years, Wireless Sensor Networks (WSN) and IoT-based solutions are increasingly adopted as monitoring systems for various industrial processes. We have seen its adaptation in the health industry and environmental monitoring, amongst many others. Consideration for their use in monitoring processes in the OGI has also increased. As discussed in 2.2.3, there are various ongoing research to improve WSN-based solutions for the OGI.

Nevertheless, our contribution is focused on the midstream sector, which is the transportation of oil and gas products. We mainly worked on a monitoring system for pipeline transportation of crude oil in the Nigerian Oil and Gas Sector. Whereas there are existing works (also discussed in 2.2.3), they are centralised. Their centralised nature makes them susceptible to Single Point of Failure (SPOF) especially considering the high rate of third-party interference and vandalism ([82–84]) in our POD (Nigeria). In addition, despite the use of such systems like SCADA by some of the major oil and gas companies in Nigeria, the detection and localising time is quite high. Between 2017 and 2021, the average detection and localisation time for one of the major oil companies in Nigeria is one day. Similarly, in 2019, another company recorded an average detection and localisation time of four days. Both records are from the data provided by the DPR Nigeria for the purpose of this research.

Therefore, we design a fault-tolerant monitoring system to detect and localise leakages and failures for crude oil transportation via pipelines in a timely manner. Some of the critical factors we consider in this design are network coverage for the communication aspect, the monitoring system’s sensitivity to leakage, detection accuracy, fault tolerance, and energy consumption.

Thus, in this chapter, we present the first part of our work: a WSN-based distributed detection and localisation of leakages in a single horizontal transmission pipeline published in [130]. In the first section 4.1, we discuss factors and choices that enable distributed leakage detection and localisation. In section 4.2, we introduce the algorithm-HyDiLLEch, followed by the simulation results in section 4.3. Finally, we conclude the chapter in section 4.4.

4.1 Collective Detection of Multifarious leakages

In this section, we introduce the techniques based (on node placement and data correlation) on which we developed a distributed leakage detection technique. The global objectives are as follows:

1. **Coverage:** Allows optimal connection between sensors.
2. **Sensitivity:** Determine multi-sized leakages, i.e. small to big-sized leaks.
3. **Fault tolerance:** Continue to detect leakage in the event of failures.
4. **Accuracy:** Ensure a highly accurate localisation of leakages
5. **Energy efficiency:** balance the energy consumption without compromising accuracy
6. **Detection time:** Implement this detection and localisation in real time assuming leakages occur linearly i.e., one after the other in every pipeline segment.

We present the node placement and data correlation strategies in the following subsections. Other objectives are discussed in the next section.

4.1.1 Node Placement

Node placement strategies significantly impact the efficiency of a Leakage Detection Monitoring System (LDMS). Existing strategies for node placement include those based on the maximum transmission capacity of sensor nodes [73], determining the shortest distance between the node and event [67,68] or most commonly, placement at the key junction of the pipelines as shown in Fig. 4.1 using the typical sensors. Some of the disadvantages of the existing methods, for example, in the case of maximum transmission range, are; to communicate between two nodes, the sensors always expend the maximum energy given the distance between them, thereby increasing the average energy consumed.

Moreover, failure of one of the sensor nodes, e.g. an intermediate node, can interrupt the transmission as new neighbouring node(s) will cease to be in each other's range. In addition, using the minimum distance between the event and sensor may not be practical because leakages and other failure events on the pipeline are stochastic in nature, i.e. they cannot be predetermined. Finally, consider the sensor placement in Fig. 4.1, assuming it represents long-haul transmission pipeline networks. Although it is economical to have such placement, its drawbacks include its insensitivity to small leaks, given the length of each pipeline segment and vulnerability to node failures.

For example, it is impossible for pipeline segments II, III, IV and V or IV, VII to detect or localise leakages if the sensors in those junctions fail. Putting it in context, that is, between 14% or 57% of the pipeline network goes out of coverage due to a single node failure.

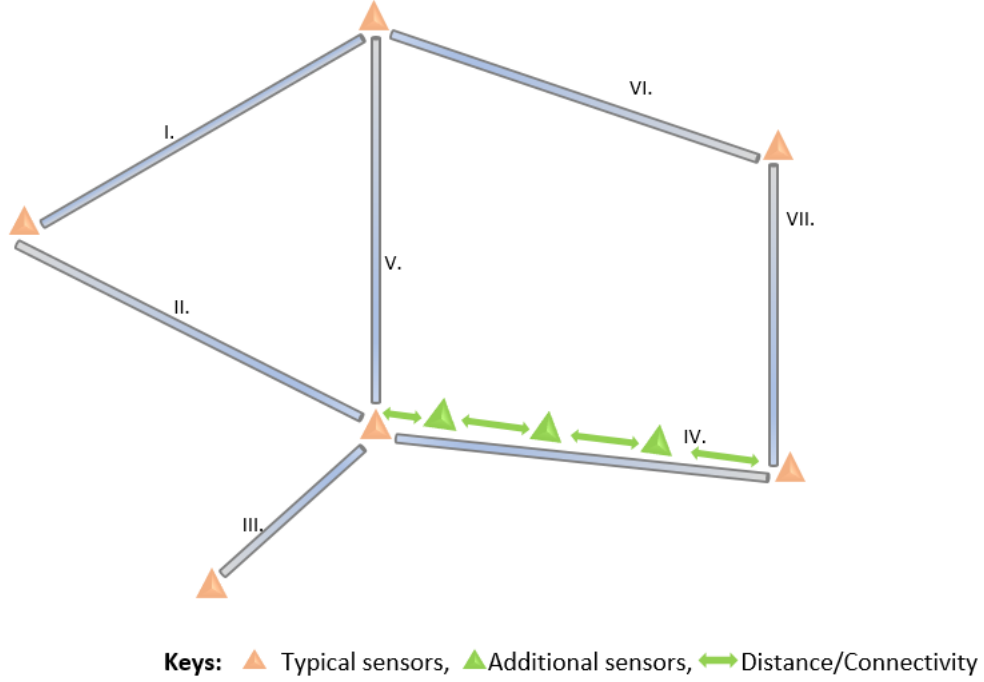


Figure 4.1: Sensor placement on long transmission crude oil pipelines

Hence, we propose deploying several sensors along a pipeline segment contrary to the deployment at key junctions. However, in our work, we recommend a new node placement strategy based on fluid propagation properties -NPW- when leakage occurs in a pipeline to determine the optimal distance between sensors. Based on the NPW phenomenon and the PG properties discussed in subsection 2.1.2, sensors can be placed at the shortest distance from the source. Specifically, where events like small, medium and big-sized leakages are detectable by examining the amplitudes of the NPW for various-sized leaks and the corresponding attenuation. To find such distance D on a pipeline of Lkm , we apply the following constraints:

- The distance (D) between sensor nodes must be less than half the maximum communication range of the sensor (S_{cr}) to secure interconnectivity and data sharing between sensors

$$D < S_{cr}/2 \quad (4.1)$$

- The maximum detecting distance of the NPW (NPW_{mdd} to be determined experimentally) between sensors and event source is based on the amplitude of the NPW, and it must be adequately small to guarantee sensitivity to small-sized leaks.

$$0 < D < NPW_{mdd} \quad (4.2)$$

where, D is the distance obtained in Eqn. 4.1

According to Eqn. 4.2, the distance (D) between the sensors, should be less than the NPW_{mdd} between the upstream and downstream sensors surrounding a leakage. This way, the NPW travelling in both directions can be detected.

- There should be at least three sensors that can detect the NPW front from the total number of deployed sensors (N). Indeed, in an ideal case, the upstream and downstream nodes are enough to detect the arrival of the NPW front. But, to prevent SPOF and the impossibility of localising a leakage, we add a minimum redundancy of one node to provide a continuous detection and localisation in the event of node failure.

$$\forall Q, [\exists n_1, n_2, n_3 : \text{dist}(n_i, Q) < (NPW_{mdd})] \quad (4.3)$$

where, $i > 0$ and $\text{dist}(n_i, Q)$ is the distance between node i and the leak location Q .

Utilising the constraints defined above, we discuss the data correlation among the sensor nodes in the next subsection.

4.1.2 Spatial Data Correlation

We propose spatial correlation among geographically close sensors to allow distributed leakage detection and localisation. Our proposed method ensures that sensor nodes can communicate, allowing collaborative data processing to improve detection efficiency and reduce or eliminate false positives. However, it is expensive and almost impossible to have all the sensor nodes interconnected, i.e. fully meshed connection. Hence, we chose a limited multi-hop communication to achieve this collaboration among the nodes.

The benefits of a multi-hop connection compared to a single wireless link includes an increment in the network coverage, higher connectivity and reduction in transmission power [131]. It can also improve the throughput due to a higher data rate. However, linear deployment of a multi-hop communication network has its shortcomings. They include the rise in energy consumption towards the sink node as the number of hops before the sink node increases leading to unbalanced energy consumption. The number of messages per sensor also increases significantly in a linear deployment with an

increasing number of connections among the nodes. In addition, multi-hop communication in wireless networks introduces interferences that can significantly affect the efficiency and performance of the network [132]. It also introduces high communication overhead, which may be difficult to eliminate through data aggregation [131].

Therefore, when we consider a linear deployment of sensors for the WSN layer, as shown in Fig. 3.5, we limit the number of collaborations among the sensors to a maximum of two hops. This allows the minimisation of multiple interferences which results from such collaborations while optimising the energy consumption of the sensors. In our case, the number of hops is in two directions, i.e. a single-hop connection will represent one upstream neighbour and one downstream neighbour. Hence, a focal node (n_i - the node making the detection or localisation) will use information from itself and two other nodes. Two hops will represent additional information from two upstream and two downstream nodes, as illustrated in Fig. 4.2.

4.2 HYbrid DIstributed Leakage detection and Localisation tECHnique (HyDiLLEch)

HyDiLLEch, is an LDT based on the fusion of several computational fluid dynamic detection techniques, i.e. PPA, NPWM, GM introduced in 2.1.2. Each chosen LDT has its strength and weaknesses, such as the detectability of leakage, energy consumption, the accuracy of leakage localisation and others, as enumerated in Table 2.1. For example, PPA can detect small leaks, but it is inexpensive, has low maintenance and performs well under extreme conditions. The strengths of GM and NPWM techniques include their capability of detecting and localising single or multiple leakages in transient state [48]. Together, these three methods share the characteristic of non-invasive LDTs with low computational complexity, which are more easily implemented and deployed on existing infrastructure in comparison to other methods of leak detection [48].

In any case, while the aforementioned techniques incorporate several advantages, they also have drawbacks. Consider the PPA technique; it cannot be used to detect leakages in a transient state, and it can also not be used to localise leakages independently [28]. On the other hand, the NPWM relies on accurately detecting the arrival of the wavefront at the upstream and downstream sensors. Finally, GM depends on the sensor nodes' calibration and accuracy. As a result, we use a combination of these detection techniques (NPWM, PPA and GM). This combination allows us to take advantage of their strengths to enhance leakage detection and localisation accuracy without their negative impacts. This will be discussed in more detail in the following subsections. HyDiLLEch is mainly targeted at increasing the *fault tolerance*

of the LDMS, improving the *accuracy* in leakage detection, removing or minimising *false positives* all in an *energy efficient manner*. This is achieved through limited collaborations among the nodes, reduction in the sensing and multi-data analysis.

Algorithms 1 and 2 show the main structure of HyDiLLech. It is implemented with a neighbourhood of 1-hop (2 nodes) or 2-hops (4 nodes), both referred to as HyDiLLech-1 and HyDiLLech-2. The two algorithms have similar formulations, only differing in the number of nodes participating in the neighbourhood collaboration. Both algorithms are implemented on a single horizontal pipeline as illustrated in Fig. 4.2. Note that the sensors used are multispectral, i.e. can detect several fluid properties such as speed, pressure, or temperature and equipped with it's own battery. In the following subsections, we explain the detection and localisation techniques in detail.

Algorithm 1 HyDiLLech (Single-Hop)

```

1: {Init-steady state}
2: Set upper and lower PG thresholds
3: for ever do
4:   Get pressure data from neighbours
5:   Calculate local gradients  $PG_{(i-1)-(i)}$  and  $PG_{(i)-(i+1)}$ 
6:   if local PG is outside the threshold then
7:     Scan at high frequency
8:     for scanning time do
9:       Get  $P_i$ 
10:    end for
11:    if exists  $P_i$  greater than threshold, then share NPW data then
12:      Localise using GM data (equation 2.2)
13:      Localise using NPW data (equation 2.5)
14:    else
15:      No leak detected
16:    end if
17:  else
18:    No leak detected
19:  end if
20:  Sleep (duty cycle)
21: end for

```

HyDiLLech first defines the gradient thresholds (line 2 of Alg. 1). These values are calculated to prevent classical variations of the fluid propagation model and customised offline. Each sensor capture the pressure and propagate/receive the values

Algorithm 2 HyDiLLEch (Double-Hop)

```
1: {Init-steady state}
2: Set upper and lower PG thresholds
3: for ever do
4:   Get pressure data from neighbours
5:   Calculate local gradients  $PG_{(i-1)-(i)}$ ,  $PG_{(i-2)-(i)}$ ,  $PG_{(i)-(i+1)}$  and  $PG_{(i)-(i+2)}$ 
6:   if local PGs are outside the threshold then
7:     Scan at high frequency
8:     for scanning time do
9:       Get  $P_i$ 
10:    end for
11:    if exists  $P_i$  greater than threshold, then share NPW data then
12:      Localise using GM data (equation 2.2)
13:      Localise using NPW data (equation 2.5)
14:    else
15:      No leak detected
16:    end if
17:  else
18:    No leak detected
19:  end if
20:  Sleep (duty cycle)
21: end for
```

to its neighbours (line 4 of Alg. 1). Pressure gradients can then be computed to see if unexpected pressure drop happened (line 5 of Alg. 1). If so, to determine the location, several pressure values are sensed at high frequency to capture the leak wave (lines 8-10 of Alg. 1). If a wave is detected, the precise location is computed from the changes in the gradient, wave amplitude and arrival time (line 12 and 13 of Alg. 1). After the detection and localisation, the sensor goes back to sleep during a period of time (line 20 of Alg. 1). This period of time is the network duty cycle - i.e. the ratio of time the sensor is ON compared to the time it is OFF (for instance 60% duty cycle means the sensor is working 60% of time and sleeping 40% of time).

4.2.1 3-factor Leakage Detection

Leakage is detected in a multiple manner in HyDiLLech using the *predefined pressure threshold*, the *pressure gradient* and the *arrival* of the NPW front at the sensor. Each method of detection is further explained below.

Defining the pressure threshold

The algorithms begin by utilising the PPA LDT to pre-estimate the expected pressure at every sensor node location in the pipeline, as shown in Fig. 4.2. In the first part of our work, we consider a single horizontal pipeline, as shown in Fig. 4.2. Thus, to determine the expected pressure using the PPA technique, we set the elevation parameter (z) defined Eqn. 2.1. Assuming our pipeline has a total length of L kilometres, then elevation $z_{0-L} = 0$.

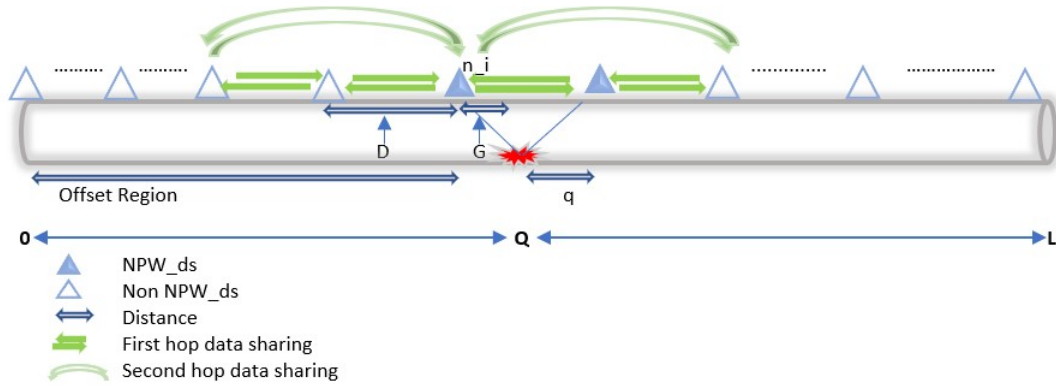


Figure 4.2: Detection and localisation of leakages

The pressure gradient helps us estimate the expected pressure at every sensor point. In our case, we can rewrite Eqn. 2.1 as follows. Hence, PG_{0-L} can be estimated as follows:

$$PG_{0-L} = \left(\frac{P_0}{\rho g} - \frac{P_L}{\rho g} \right) \frac{1}{L} \quad (4.4)$$

where PG_{0-L} is the pressure gradient for a horizontal i.e. elevation $z_{a-b} = 0$ pipeline of length L , (P_L) pressure at the outlet (P_0) is the pressure at the inlet, ρ is the fluid density and g is the gravitational force.

Once the fluid's rate of change in pressure as it travels along the pipeline in a steady state is determined, a threshold ϵ is set to account for the difference between the value read from the sensor and possible calibration error from the sensor readings. This threshold is set using the industrial permissible standard [48]. This set the basis for comparison between the reading of the sensor and the value obtained from the pre-estimation with Eqn. 4.4. A difference greater than the threshold indicates possible leakage occurrence.

Difference in the pressure gradient

To substantiate this indication, we check the two PGs that must be present in case of leakage -based on fluid dynamic properties explained in 2.1.2. These two pressure gradients are the pressure gradient formed between n_i and n_{i+1} assuming a leakage point as shown in Fig. 4.2 is at location Q . Both gradients, i.e PG_{0-Q}^{leak} and PG_{Q-L}^{leak} must be present. Then, we also ensure that the following equations are true:

$$PG_{0-Q}^{leak} \neq PG_{Q-L}^{leak} \quad (4.5)$$

$$PG_{0-Q}^{leak} < PG_{0-L} \quad (4.6)$$

$$PG_{0-L} < PG_{Q-L}^{leak} \quad (4.7)$$

Ensuring the correctness of Equations 4.5, 4.6, and 4.7 further reduces the high false alarms associated with other detection techniques.

Arrival of the negative pressure wave

As discussed in chapter 2, a leakage must necessarily generate a negative pressure wave (NPW) in the pipeline. Thus, to finalise the detection process, certain sensors among all the sensors detect this wavefront. We identify these sensors as the NPW_{ds} in Fig. 4.2 and are those sensors around the areas the pressure gradient is formed.

Note that sensors are placed in such a manner that when leakages occur (including small leakages), the sensor is able to detect them. Thus, we reduce or even eliminate false positives by coupling leakage detection to the arrival of the wavefront. In addition, this arrival for it to be considered will occur in at least two (one upstream and one downstream) of the sensor nodes. This condition is based on the property of the generated wave when leakage occurs, i.e. the wave travelling in opposite directions from the point of leakage.

Ultimately, we confirm the actual presence of leakage following these three factors. The localisation of the leakage is estimated using a similar approach as discussed in the following subsection.

4.2.2 2-factors Leakage Localisation

To localise the leakage, we iterate over the sensors on the pipeline by combining the data (PG, NPW) to narrow down the region of the leak, i.e. the region where the NPW is detectable and where the change in gradient occurred. This gives us an offset region calculated as $-D \times i$ where i is the index location of the node. Assuming a focal node n_i as shown in Fig. 4.2 receiving data from nodes n_{i-1} and n_{i+1} in a single-hop neighbours and additional data from nodes n_{i-2} and n_{i+2} from double-hop neighbors, the leakage can be localised in two ways as follows:

$$\overline{Q}_{dpNPW} = (D \times i) + (D - q) \quad (4.8)$$

and

$$\overline{Q}_{dpGM} = (D \times i) + G \quad (4.9)$$

where \overline{Q}_{dpNPW} is the leakage location based on the *partial information* on the NPW and estimated by Eqn. 2.5, \overline{Q}_{dpGM} is the leakage location based on the *partial information* on the pressure gradient, and G is the distance based on the gradient calculation.

Both equations 4.8 and 4.9 provide absolute leakage position - the first estimated position when receiving the first leakage front-wave. According to the amplitude and speed of these wave, the absolute position can be refined through the pressure gradients and Eqn. 2.2 or the time of arrival of the NPW front as follows:

$$\overline{Q}_{dpNPW} = (D \times i) + \sum_{n=1}^k (c \times \delta t_n) \div k \quad (4.10)$$

where \overline{Q}_{dpNPW} is an absolute leakage location, n is the n -th front end of the leakage wave, c is the negative wave speed estimated by Eqn 2.3

of the n -th front end of the leakage wave, δt_n the difference in arrival time of the n -th front end of the leakage wave at the upstream and downstream nodes, and k is the number of received front-end wave.

HyDiLLEch is expected to eliminate the SPOF problem related to centralised systems through sensor-based detection and localisation in more than one node as a result of the various techniques put into consideration in its development. It is also expected to have relatively high detection and localisation accuracy with minimised energy consumption. The energy consumed is reduced by utilising only a fraction of nodes in the localisation process. Thus, we discuss and analyse the simulation results in the next subsection.

4.3 Simulation Results

In this work, implementation is carried out with several goals. The first is determining the distance at which nodes can be placed for sensitive leakage detection. In addition to that, we also carried out the implementation of some classical LDTs (NPWM and GM) to enable comparative analysis with our proposed distributed LDT (HyDiLLEch). We evaluate their accuracy in leakage detection, susceptibility to SPOF and, in addition, communication efficiency in terms of energy consumption and communication overhead. We used NS3 to simulate crude oil propagation and leakages to achieve this. To cover the typical distance of a Lora-based communication while implementing the mesh connectivity between the nodes, we used the WiFi protocol in NS3 with an increased transmission power. The energy consumption of both LoRa and WiFi differs by a constant factor after a distance of approximately 350m in a linear deployment i.e., node placement with a constant distance between the nodes [133]. Given the similarity in node deployment in our work with a constant distance (D) greater than this threshold, we assume the corresponding energy consumption using WiFi is also comparable by a constant factor to LoRa. To determine the distance between the nodes, we first carried out a preliminary experiments and the results are discussed in the next subsection.

4.3.1 Node placement

The main focus of this phase of the work is to determine the optimal distance D between uniformly distributed sensor nodes on the pipeline. This partially fulfils the objectives of this contribution, i.e. ensuring the sensitivity of HyDiLLEch to small and large leaks. Thus, we carried out an analysis of the impact of several amplitudes on the detectability of leakages using predefined pressure thresholds for PPA and the amplitude of NPW for NPWM. PPA technique works in the detection process

in Fig. 4.3. We conducted the simulations using $0.5kpsi$, $5kpsi$ and $20kpsi$, each representing small, medium and large-sized leaks at various distances between the sensors. Obtained results show that for the PPA technique, all sensor nodes along the pipeline can eventually detect the leakage based on the pre-estimated pressure and a pre-defined threshold value allowing a gradient-based detection. However, the NPWM shows maximum detectable distance NPW_{mdd} of approximately $2500m$ for the highest amplitude of $20kpsi$. In the case of the least examined amplitude $0.5kpsi$, the NPW_{mdd} is approximately $1500m$.

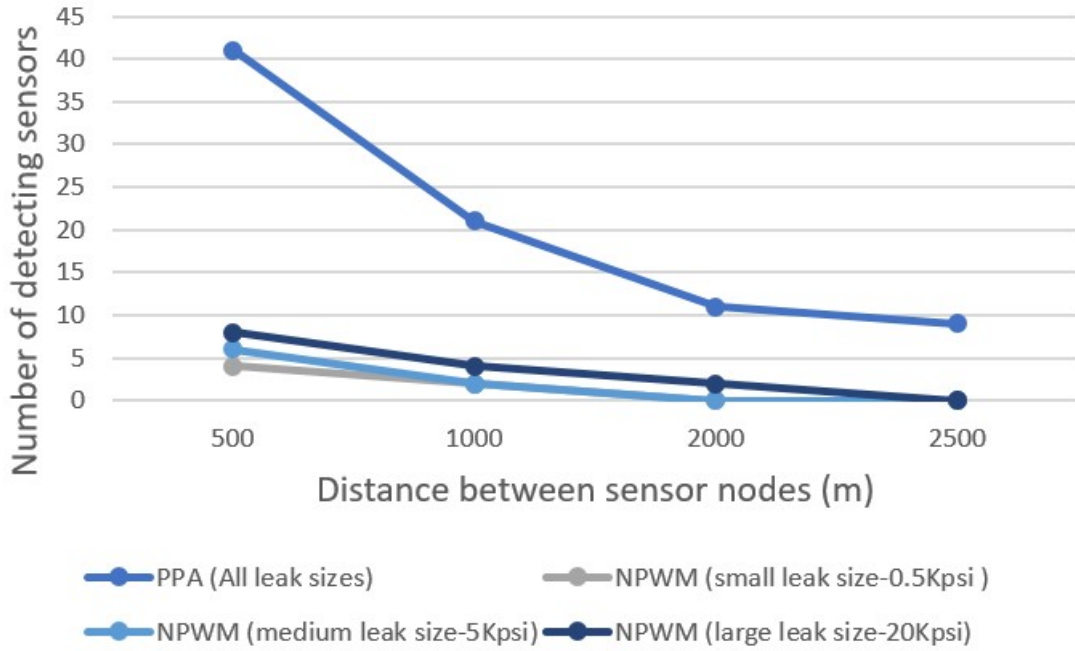


Figure 4.3: Leakage detectability by number of sensors and distance

As depicted in Fig. 4.3, the NPWM, -i.e. detection technique based on NPW- is not very efficient in long-distance leakage detection. This is shown by the number of sensors over which the amplitude of the NPW is detectable, i.e. the number of sensor nodes that can detect the NPW decreases as the distance from the leakage point increases. However, the smallest leak tested is detectable beyond $1000m$ and by multiple sensors. When all the leak sizes are considered, we can see from the figure that, at a distance of $1000m$ away from the leakage point, the NPW is still detectable by an average of three sensors. Also, the constraint defined by Eqn. 4.3 is satisfied at this distance with an average of three sensors detecting the NPW. Additionally, all other conditions listed in section 4.1 are satisfied. Thus, for the sake of this

work, we choose D equals 1000m as the distance between the sensors used for all the simulations.

With the optimal distance determined, we implemented HyDiLLEch on a horizontal long-haul transmission pipeline for a single-phase laminar flow. The pressure data generated by the sensors follows the Bernoulli equation defined by Eqn. 2.1. Also, the fluid propagation parametrisation results from the collaboration with the DEEP laboratory ¹- working on environmental management. Additionally, the properties taken into consideration for this simulation include industry-defined specifications such as the operational velocity, type of crude oil, and length and material of the pipeline, to mention a few. The data for defining these properties are from existing pipeline networks and sourced from Nigeria's Department of Petroleum Resources (DPR), including preliminary results from the first simulation works. Table 4.1 outlines the simulation parameters and crude oil properties.

Material	Carbon steel
Pipeline Length (L)	20km
Wall thickness (w)	0.323m
Inside diameter (d)	0.61m
Height/elevation (z)	0m
Oil kinetic viscosity	2.90mm ² /s
Temperature	50°C
Oil density (ρ)	837kg/m ³
Inlet pressure (P_0)	1000psi
Reynolds no (Re)	1950
Velocity (V)	2m/s
Molecular Mass (m)	229
Oil elasticity (K)	1.85 × 10 ⁵ psi
Carbon steel elasticity (Y)	3 × 10 ⁶ psi
Gravitational force (g)	9.81m/s ²
Constant (e)	2.718
Coefficient of friction (λ)	0.033
Wave speed (c)	14.1m/s

Table 4.1: Pipeline and oil characteristics

Number of sensors	21
Number of gateways	1
PHY/MAC model	802.11ax Ad hoc
Transmit power	80dBm
Transmit distance	20km
Error model	YANS
Propagation Loss	Log-distance
Path Loss (L_0)	46.67dB
Reference distance (d_0)	1m
Path-Loss Exponent (σ)	3.0
Packet size	32bytes
Data rate	1Kbps
Distance between sensors	1Km
Duty cycle	70%

Table 4.2: Network simulation parameters

As a preliminary work, all tests are conducted in ideal conditions, i.e. no communication or node failures. In the following subsection, we discuss the results.

4.3.2 Detection and Localisation

To conduct the detection and localisation test, a random selection of sample leakage points was drawn along the length of the pipeline. The sample points have a confidence interval of approximately 9880m – 14523m and a confidence level of 98%. The localisation accuracy for each LDT was calculated formerly calculated based on percentage as in [130]. However, we noted that this form of presentation does not

¹deep.insa-lyon.fr

show the accuracy with respect to the distance in meters. Thus, in this document, we present the localisation accuracy as distance in meters from the actual leakage position.

Following, we discuss the result obtained using classical LDT such as NPWM and GM in comparison with HyDiLLEch.

Classical Approach

For the centralised approach, we implemented two LDTs (NPWM and GM) using all the information collected at the gateway. The result obtained from the simulation is shown in Fig. 4.4. The GM present an average localisation accuracy error of approximately 227m with a high variance across the tested leakage points. NPWM, on the other hand, shows a more consistent localisation accuracy with an error in distance of about 2m and almost insignificant variance. However, this level of accuracy is only achievable with a high sampling rate of all nodes resulting in high energy consumption.

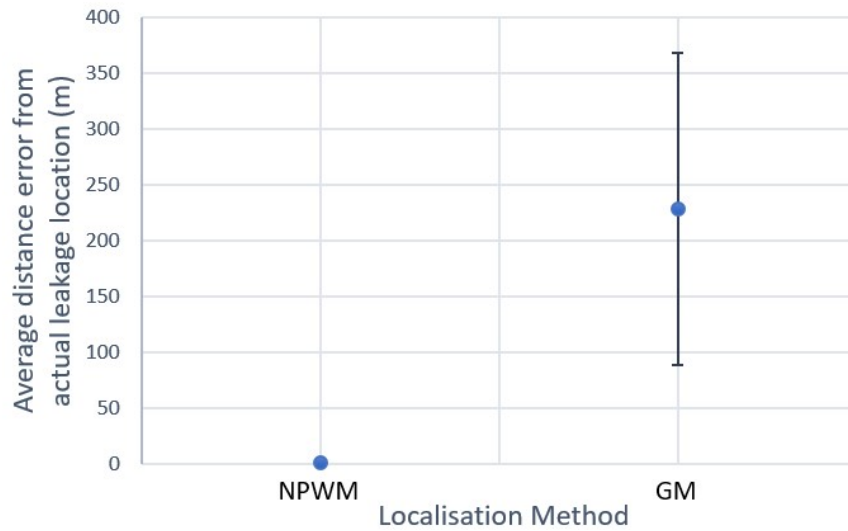


Figure 4.4: NPWM and GM

Although this simulation demonstrates the efficiency of both LDTs in the localisation of leakages, the centralised approach to the detection and localisation makes it vulnerable to SPOF and less robust to other types of failures, i.e. communication failure and vandalism. In the following subsection, we discuss the advantages of HyDiLLEch relative to this drawback by considering the increment in the number of nodes detecting and localising leakages (NDL).

HyDiLLEch

For the implementation of the distributed approach, both versions (single and double-hop) of HyDiLLEch were separately considered. We break down the results based on NDL (by their distance from the leakage point) and also the principal information (dpNPW and dpGM) considered in the localisation process. As shown in Fig. 4.5, the results for the single-hop version (HyDiLLEch-1) show an increment in the number of sensor nodes detecting and localising leakages (NDL) from one to four in comparison with the centralised approach. While the centralised systems does localisation in a single NDL, HyDiLLEch utilises multiple nodes surrounding the leakage area for this purpose, thereby increasing its fault tolerance. The NDLs which are represented as n_{1-4} in the figure, results from the spatial correlation of data from neighbouring sensors. This is in addition to partial information such as the sensor geolocation, the pressure gradient (dpGM), and the time of arrival of the NPW front (dpNPW) at each NDL. In HyDiLLEch-1, we refine the dpGM technique active on all nodes and active dpNPW at high frequency on the nodes whose gradient information differs from one of the neighbours, i.e. the nodes closest to the leakage points. Thus, the average error from the actual leakage location is about 2m for the best NDLs (n_2 and n_3) in the case of dpNPW and about 33m for dpGM.

As shown in the figure, nodes n_2 and n_3 maintaining the highest accuracy are the ones closest to the leakage point. The nodes farthest (n_1 and n_4) from the leakage point have an average of about 298m in the case of dpNPW. We notice that, although the detection is possible, both nodes are too far away from the leakage to precisely evaluate the leakage location. Note that to use dpGM for localisation, gradient information from at least one upstream and one downstream node of the leakage point is required. Hence, in this case, only dpNPW can be used for localisation at the extremities (n_1 and n_4). However, these values are kept to be potentially filtered at the upper layer as outliers. Also, in the case of failures, they still give a precision error averaging between 230m to 367m on a 20km pipeline that is considered.

Similarly, Fig. 4.6 shows the result obtained for the double-hop version of HyDiLLEch (HyDiLLEch-2). In this version, the NDL increased from 1 to 6 and is also represented as n_{1-6} . The differing number from the single version is a result of the increment in the number of nodes participating in the neighbourhood collaboration. Sensor nodes that are physically-close to the leak - n_2, n_3, n_4, n_5 - also maintains the highest accuracy. The increment in spatially correlated data reduced the error for both dpNPW at dpGM. For dpNPW, we have an absolute detection and average of 32m error in distance across the four nodes for dpGM, with lesser variation. In addition, the average error in distance at the extreme node is slightly reduced to 288m compared to HyDiLLEch-1.

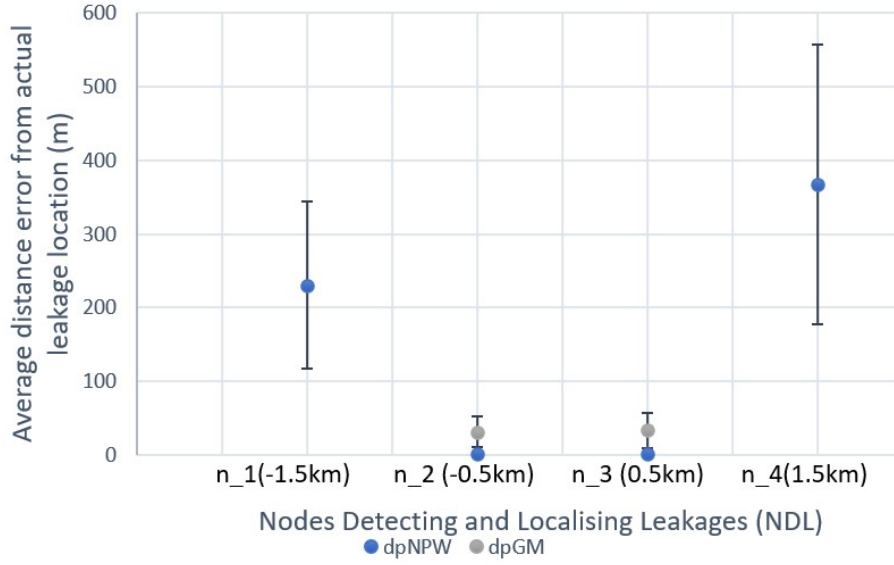


Figure 4.5: HyDiLLECh-1 Average localisation accuracy by NDLs

Note that this hybrid technique is kept loosely coupled to ensure the robustness of detection in addition to enhancing the fault tolerance of the system. This is of particular interest in cases where performance-influencing factors such as communication failure, node failure and others are particularly detrimental to the partial information utilised concerning each localisation technique.

Finally, both versions (HyDiLLECh-1 and HyDiLLECh-2) showed a significant improvement in NDL (upto 6) compared to the single NDL obtained in the centralised versions. Also, the range of detection and localisation is increased to about 3000m to 5000m for both HyDiLLECh-1 and HyDiLLECh-2. Although the nodes at the extremities in both cases have the worst localisation accuracy, in the case of failure, results from these nodes can still be utilised.

4.3.3 Communication Efficiency

The three LDTs are also compared based on their communication efficiency. For this, we put into consideration communication costs such as use number of packets, the energy consumed as a result of sampling and the radio of the net devices. In the next subsections, we discuss the obtained results.

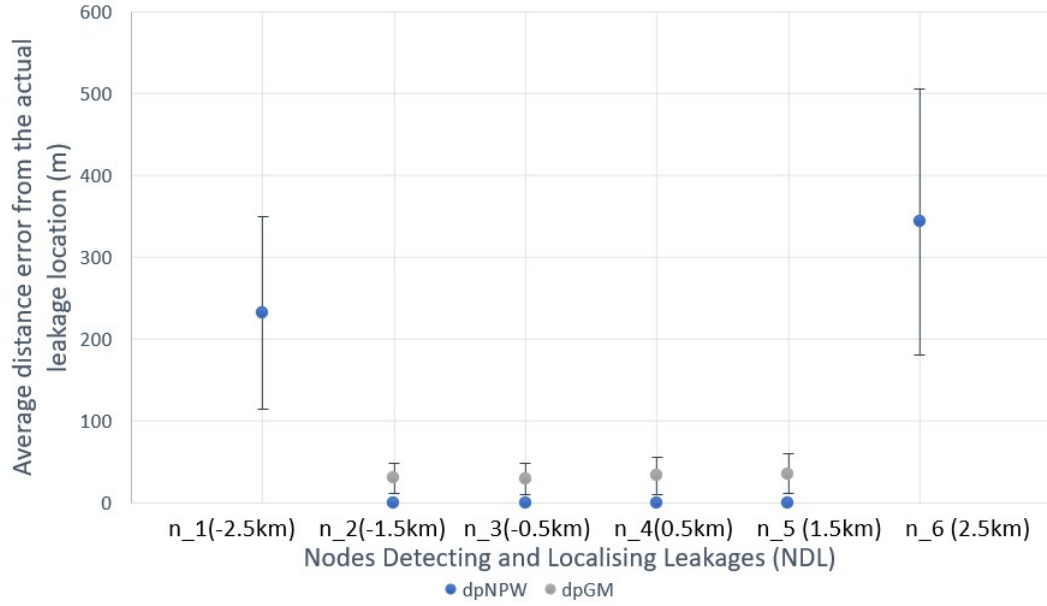


Figure 4.6: HyDiLLEch-2 Average localisation accuracy by NDLs

Communication Overhead

The number of exchanged packets (the total number of packets used for leakage localisation) is shown in Fig. 4.7. For NPWM and GM, this information is shared between the sensor nodes and the gateway, while in the case of HyDiLLEch-1 and HyDiLLEch-2, it represents information shared amongst the neighbouring sensors. Results depict the lowest number of packets exchanged for the GM compared to the NPWM and HyDiLLEch-1, and HyDiLLEch-2. The increment in the two versions of HyDiLLEch results from the data correlation among the neighbours. As such, there is 50% communication overhead from HyDiLLEch-1 in comparison to the GM and 60% in the case of HyDiLLEch-2. On the other hand, both versions show better communication efficiency than the other classical LDT-NPWM. In addition to that, it has a higher NDL to the tune of 4 and 6 increment for HyDiLLEch-1 and HyDiLLEch-2, respectively.

Energy Consumption

The evaluation of the energy consumption for each LDT is based on the sampling rate utilised by the sensor and the radio energy consumption of the network duty cycle. For the energy model we consider, the sum of energy utilised in different states such as, sensing, reception, transmission, sleep, and idle states based on the WiFi energy

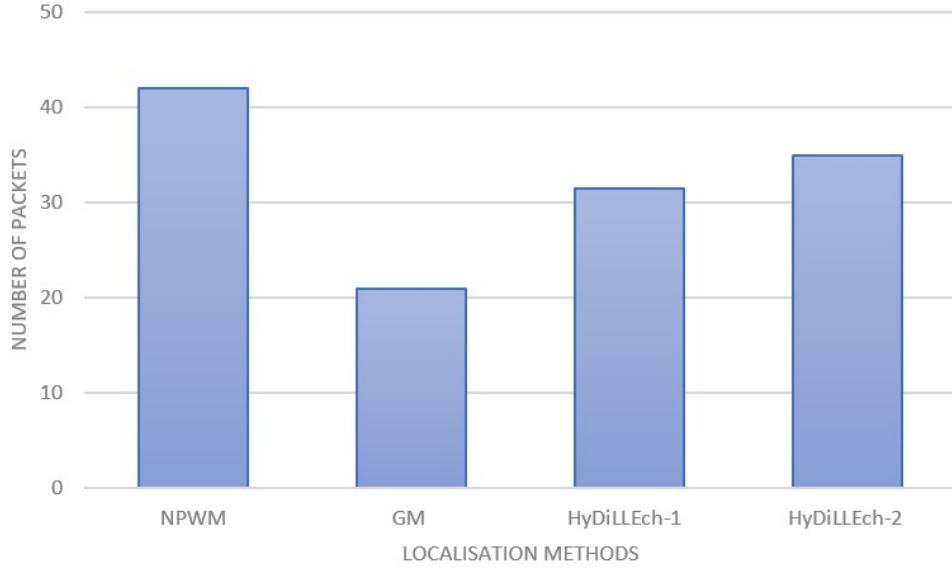


Figure 4.7: Communication overhead by the number of packets

model in NS3. In Fig. 4.8, we represent the energy consumption of the sensors per sensor. NPWM has the highest energy consumption as a result of its high sampling rate in comparison to GM and also HyDiLLEch. The combination of information utilised by HyDiLLEch allows first to determine the offset region relative to each focal sensor before the enhancement of location accuracy with either the expected time of arrival of the NPW or the gradient. Therefore, less than 2% of NPWM sampling rate is required to obtain nearly the same accuracy as NPWM while maintaining similar energy consumption of the GM.

In the case of radio energy consumption, the rate of change of the energy consumed at each duty cycle by the net device was analysed. As shown in Fig. 4.9, the most significant change is recorded in the first and second duty cycles. This change results from the difference in connections requirement for the participating nodes of each LDT. The subsequent cycle after establishing the connections shows convergence in energy consumption.

On the other hand, Fig. 4.10 shows the cumulative energy consumption of all LDTs. In comparison to NPWM, there is a reduction in total energy consumed by 86% and 83% for HyDiLLEch version 1 and 2, respectively, in the first duty cycle. An increment by 6 – 7% compared to GM per NDL is also noticed. However, both versions of HyDiLLEch show a similar growth rate of linear increment to GM as the cycle increases.

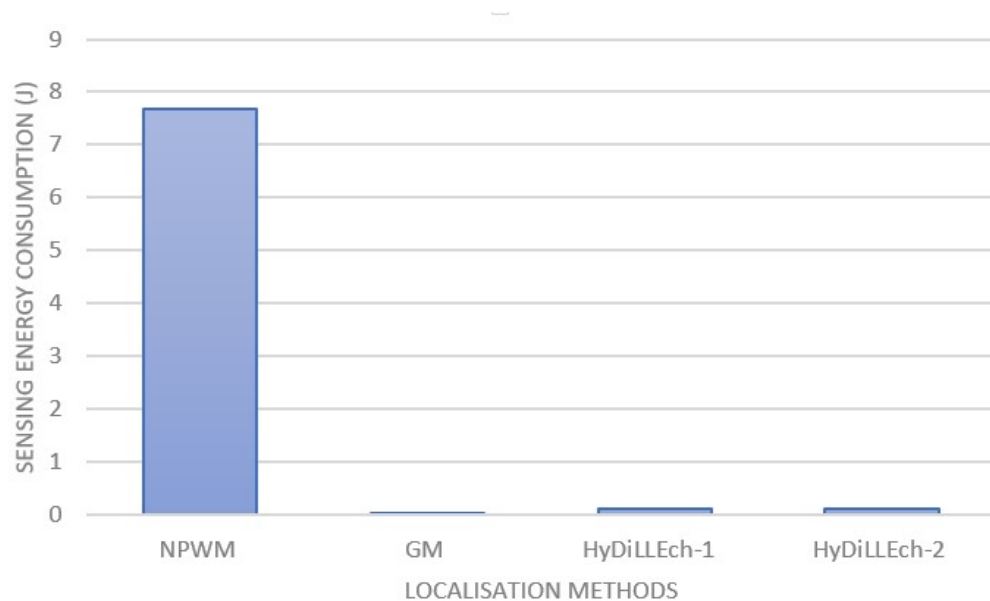


Figure 4.8: Sampling energy consumption of the sensors

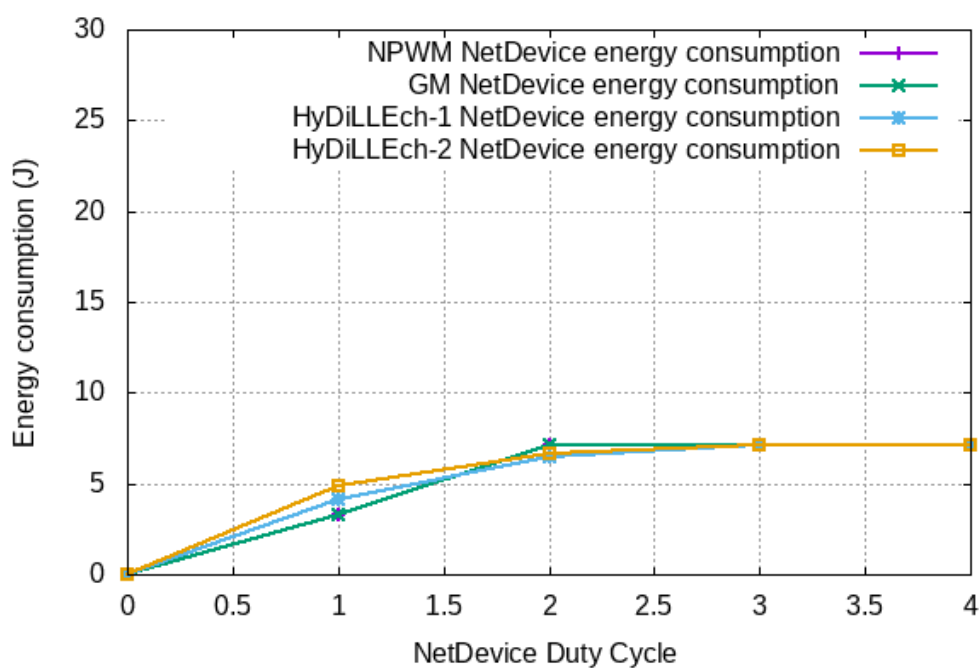


Figure 4.9: Radio energy consumption

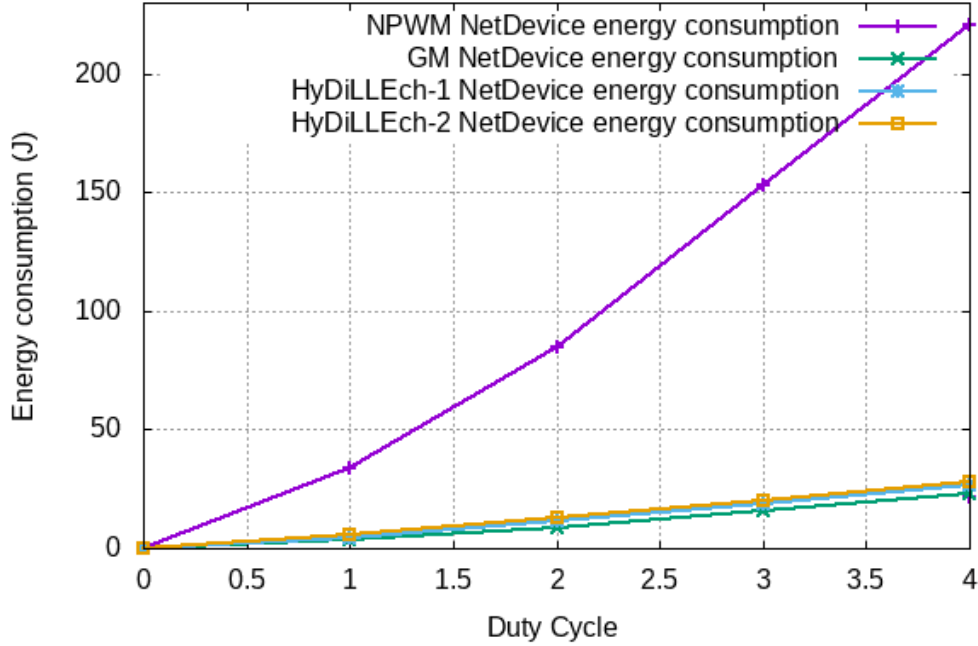


Figure 4.10: Cumulative energy consumption

4.4 Conclusion

In this chapter, we introduced a new LDT -HyDiLLEch- based on several detection techniques using a unique node placement strategy that allows distributed detection and localisation of small-sized to big-sized leakages. To obtain the optimal node placement and carry out a comparative analysis of the efficiency of HyDiLLEch, we implemented a couple of existing LDTs, such as the NPWM, GM and PPA. For the node placement, we made a choice of $1000m$ between the sensors to allow for multiple-sized leakage detection. The efficiency of the LDTs was compared in terms of the number of NDLs, energy consumption and communication overhead. Compared to currently used detecting and localising methods presented in chapter 2 and the data provided by DPR for oil and gas companies in Nigeria, - current detection and localisation time is in order of days. HyDiLLEch with single or double-hops detecting and single or double-hops notifying works in seconds.

Obtained results from the simulations show that HyDiLLEch in both the single and double-hops data sharing increased the number of NDL by four to six times, respectively, compared to the centralised LDTs. This increment in NDLs eliminates the SPOF problem related to the classical LDTs. In addition, HyDiLLEch showed high communication efficiency with minimal increment in the communication overhead

compared with the other LDTs. Due to the low sampling rate utilised in HyDiLLEch, the energy consumption at the network level was maintained despite the spatial data correlation. However, in this part of the work, we have not included any failure. Thus, in the next chapter, we introduce various failures, such as communication failures.

Chapter 5

A Game Theory Approach to Data and Service Management for Crude Oil Pipelines Monitoring Systems

The operations across all oil and gas sectors present terabytes of data. According to Slaughter *et al.*, [27], every 150,000 miles in the midstream sector, i.e. the transportation of crude oil, produces about ten terabytes of data. Analytic operations on such data can considerably enhance the processes of the three sectors. While these insights are well advanced [134, 135] in the upstream and downstream sectors, i.e. the exploration and distribution activities, respectively, the midstream sector has been mainly left unexploited. However, efficient management of the data produced from the midstream sector can significantly reduce operational downtime. In addition, analytics of the data can enable the timely detection of leakages in the pipeline and improve their localisation through the use of various services. Current works focus on detecting leakages in the midstream sectors using legacy and recent Leakage Detection Monitoring Systems (LDMS).

Thus in this chapter, we present our second contribution published in [136] with the following objectives:

1. To carry out a comprehensive data analysis of historical incidents of the Nigerian National Pipeline Corporation (NNPC) pipeline network.
2. Based on the analysis, create a regionalised Markov Decision Process (MDP) to ensure similar performance across the defined regions with optimised energy consumption
3. Solve the model to determine the optimal strategy for each region.

In the first section, we present the background information by introducing the NNPC pipeline network and defining the problem to be solved. This step is followed by the formal definition and detailing of the model in section II. In section III, we discuss the implementation methods and results obtained. Finally, we conclude the chapter in section IV.

5.1 Failures in pipeline transportation and modelling process

Erosion, corrosion, vandalism, equipment failure, and network failure, to mention a few are some of the causes of failures in pipeline transportation of crude oil. Existing monitoring systems only focus on the accurate and timely detection of leakages in the pipeline network. However, the monitoring systems are susceptible to third-party interference, which is one of the leading causes of failure in this mode of transporting crude oil. We aim to circumvent this problem through *efficient* data and service management that allows *continuous detection* of leakages in the presence of such failures (in both pipelines and the monitoring system). The following subsections introduce the pipeline network and its failure incidents and define the modelling process's layout.

5.1.1 Introduction to the Environment and Problem Definition

As discussed in the introductory chapter, the midstream sector has multiple modes of crude oil transportation. However, our work focuses on the pipeline transportation of crude oil in Nigeria through the NNPC pipeline network shown in Fig. 5.1. The NNPC pipeline network spans several thousand kilometres and is divided into five areas based on their geographical location. The five areas include Kaduna, PortHarcourt, Warri, Mosimi and Gombe. Each area is characterised by different failures, varying significantly from type to frequency. In Fig. 5.2, we present a snippet of the historical data by showing the rate of incidents across the areas for five years. The data presented shows a remarkable difference in incident rates in time and location. Contributing factors to this difference include weather conditions, festivities, and the area's proximity to the border, to mention a few.

Therefore, to model this problem, we consider two principal components, i.e. *failure causing elements* and *the monitoring techniques*. Several options exist in the optimal modelling of this problem, such as Byzantine failure- a common failure modelling in distributed systems. Other approaches include a game theoretical approach such as multi or single-player games. Following, we will discuss each approach in application to the problem.

Byzantine failure model

A Byzantine failure is a failure model in distributed system theory characterised by partial information. In this environment, the correct state of a failing system com-

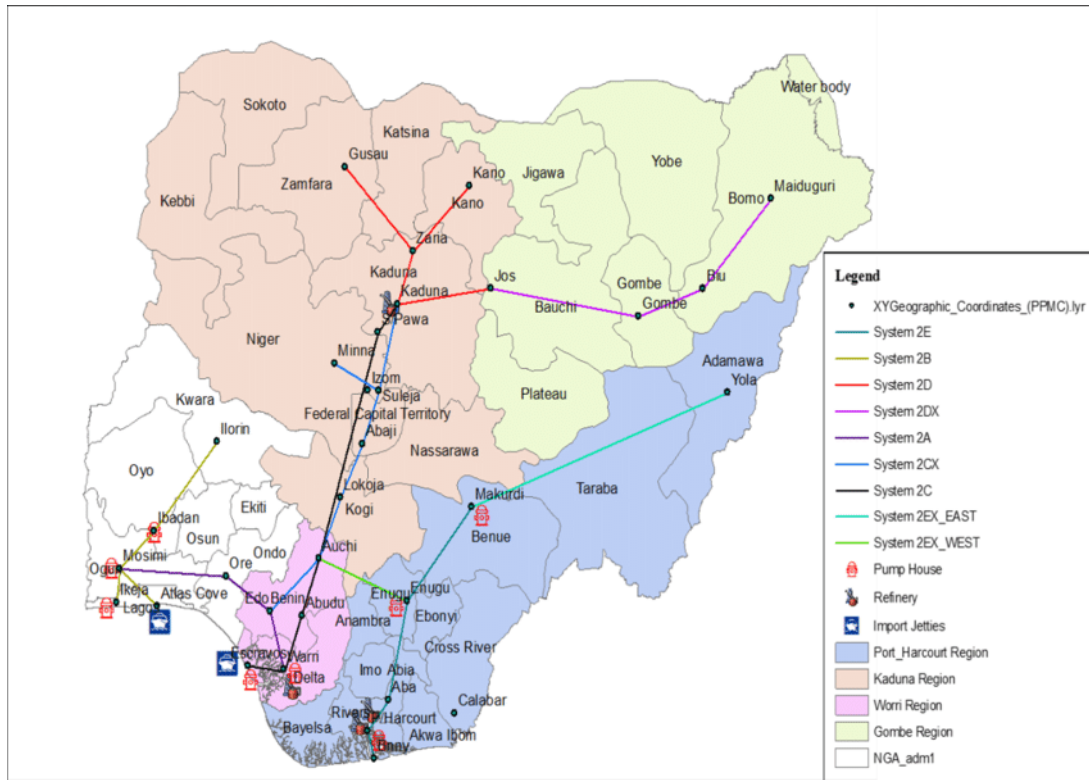


Figure 5.1: Nigerian National Petroleum Corporation pipeline network [7]

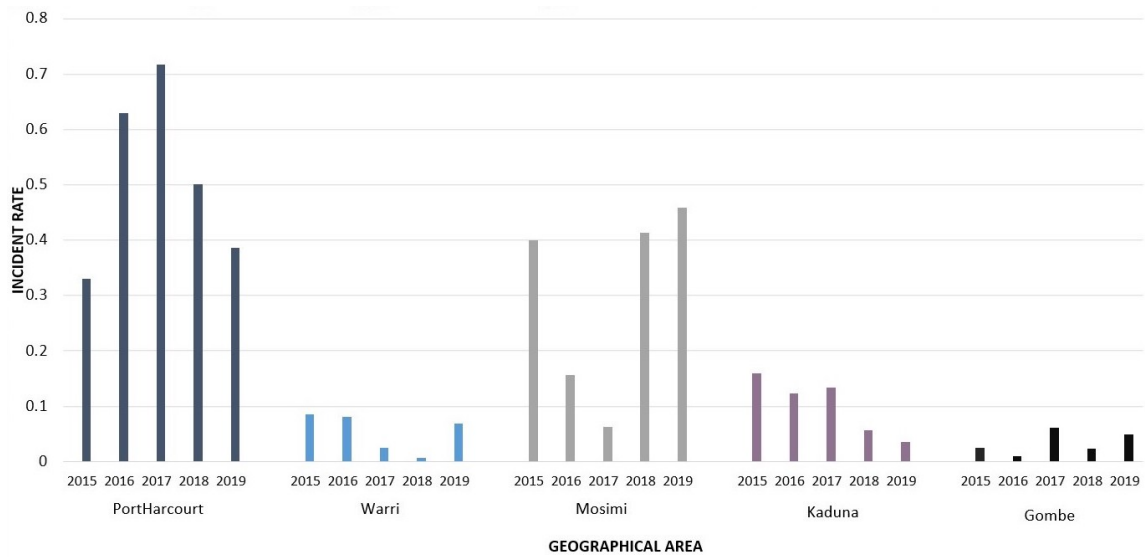


Figure 5.2: 5 year regionalised pipeline incidents [8]

ponent may not always be detectable. Hence, to avoid waste of resources or system suboptimality, a strategy is usually determined through a consensus of the non-faulty components. This approach is called the Byzantine fault tolerance (BFT), allowing a majority of the system to ensure a fault-tolerant system through a consensus. This approach to improving or ensuring fault tolerance in a system has been adopted for many use cases such as the blockchain, IoT and others [137–139]. Hence, we could apply the BFT in this system to ensure resiliency to failure.

Nonetheless, considering this problem, a Byzantine failure may result in a best-effort approach given the historical data presented in Fig. 5.2. In such a case, we will adopt the working method of a BFT where the majority of the system component must have an agreement. Suppose each area (or the nodes in each area) is considered part of the system. In that case, the majority of the areas (at least three out of the five geographical areas presented in the figure) must come to a consensus on a strategy to apply globally in the system. That is, if we consider sensors from the worst affected areas, such as PortHarcourt, Mosimi and Kaduna, then we will adopt a strategy that results in a waste of resources in areas like Warri and Gombe. Conversely, when considering the least affected areas, we will use a suboptimal strategy in the high incident rate areas.

Therefore, it is imperative to apply an approach such as a game-based approach that allows continuous and dynamic interaction with the environment and can guarantee optimal solutions in all affected areas.

A multiplayer model

System failures such as ours can be modelled as a multiplayer game. Considering the problem, on one hand, several *casuses of failures* in pipeline transportation of crude oil. On the other hand, also exists multiple *monitoring techniques* proposed as solutions to this problem. When we put this information into consideration, we could model the problem as a non-cooperative two-player game. In this case, we choose the non-cooperative game because both systems take adversarial actions. In this setting, player 1 could represent the failure-causing components while player 2 represents the monitoring system. Since our goal is to provide a fault-tolerant monitoring system against a partially stochastic opponent; therefore, we can apply the maximin strategy for player 2 (the monitoring system). The maximin strategy of a player is the strategy that maximises the player's worst-case payoff, i.e. the security level of the player [140] or the guaranteed minimum payoff. For example, let us denote a player's security level as \underline{Z}_i ; then, for player 2, the security level is defined in the following equation:

$$\underline{Z}_2 = \max_{A_2} \min_{A_1} r_2(A_2, A_1) \quad (5.1)$$

Equation 5.1 above is used to find the policy that maximises player 2's security level by taking action(s)- A_2 - and minimises the effect of action(s)- A_1 - played by player 1, i.e. the saddle point of the two players.

While pessimistic, this approach can guarantee leakage detection in the event of any type of failure on the basis of the Nash -i.e. the opponent never changes its strategy.

Nevertheless, we must define the utility or payoff of at least one player to consider this a two-player game using the maximin strategy. In the case of player 2, this could be defined as the ratio of leak detection to the actual number of leaks over a predefined period. On the contrary, this definition does not translate to a utility for player 1. In the case of player 1, most failures result from natural occurrences. Therefore, the effects of these failures cannot be considered a utility for the player. In addition, we are more interested in how the actions of player 1 affect the utility of player 2.

A single-player game

Given the discussions from the previous subsections 5.1.1 and 5.1.1, it is more rational to model this problem as a game against nature, i.e. a one-player game. This is also made practical given the historical data (Fig. 5.2) from nature, i.e. the pipeline environment that gives insight into the failure tendencies of the environment.

A one-player game can be actualised using machine learning techniques such as reinforcement learning. This technique differs from other machine learning techniques. For instance, supervised learning deals with training a set of labelled data. The objective of such training is correctly identifying or categorising objects or situations by extrapolating from the trained data set. Unsupervised learning, on the other hand, deals with the identification of structures otherwise unknown in a set of unlabelled data.

However, unlike supervised and unsupervised machine learning techniques, reinforcement learning deals with interactive problems to achieve a goal given a set of actions and continuous feedback from the environment using an MDP [141].

Thus in subsection 5.1.2, we discuss in detail the environment enabling such a decision process.

5.1.2 The Environment Setting

The setting of a game against nature enables us to maximise the utilities as we define for player 2 (the monitoring system) purely in response to the failures in our environment. Given the data presented in Fig. 5.2 showing a large diversity in incident rate from one area to the other, considering a global approach to our problem will be unnecessarily costly. Thus, we propose a local solution for each area by broadly categorising them into different *logical regions* i.e. $R = \{r_0, r_1, r_2\}$. This categorisation aims to maximise each region's utility without the cost of a globalised solution.

Following an empirical study, we define a failure rate-based threshold on which the region of the area is determined as follows: Low failure rate (r_0), average failure rate (r_1) and high failure rate (r_2).

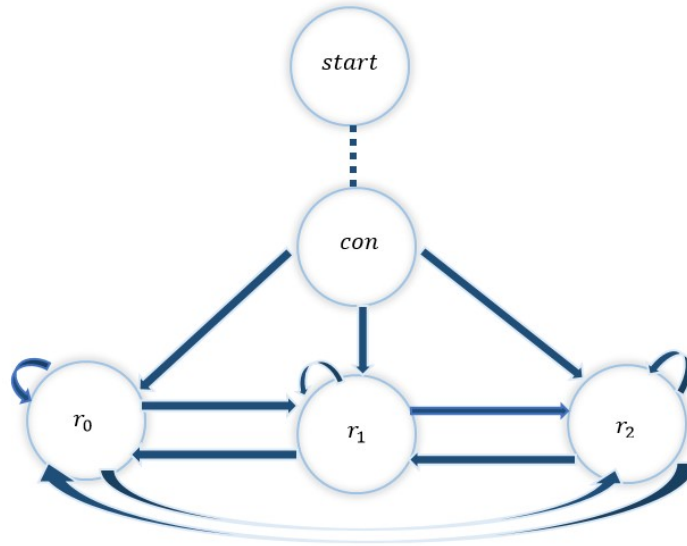


Figure 5.3: State transition Diagram

The regionalisation of the areas enables the practical implementation of data-driven strategies tailored to each region. Representing the area(s) with the least number of incidents, the area(s) with an incidents rate (IR) of 0 to 5% belongs to region r_0 . The area(s) with IR from 6% to 20% is in region r_1 , and the area(s) with IR of 21% and above falls into the critical region r_2 . Additionally, each area shows a low probability of transitioning from one region to another. For example, the PortHarcourt area consistently remains in r_2 for the period considered. On the other hand, both the Gombe and Kaduna areas are in the region r_0 and r_1 80% of the time, only changing in the year 2017 and 2019, respectively. While changing

regions approximately 40% of the timeline, Mosimi and Warri can still be considered relatively stable. Given this low transition probability, we depict each area's state in broad representation as shown in Fig. 5.3- with start and con, the non-failure states.

Further details on the state's definition will be given in the following section. The following section presents the system design and modelling based on this environment's settings.

5.2 R-MDP: A Data-Driven Regionalised Markov Decision Process

Following the definition of the regions, we aim to provide efficient and fault-tolerant data and service management using a *two-stage* approach. The first stage is to find the best detection policies in terms of convergence and optimal detection accuracy using a data-based Markov Decision Process (MDP). We build the second stage on the results of the first stage to find the policy that minimises the processes' energy consumption. Thus, we propose our objective function summarised as follows:

$$\min_{(\pi^*, V^*)} V_e(\pi^*, V^*) \quad (5.2)$$

where π^* is the optimal policy, V^* is a value function following the policy π^* , and V_e is the value of the minimum energy consumption of the node with the optimal value function

The value function obtained in each region allows us to measure and compare how good (reward and energy consumption) an applied policy is in that region. The performance measure is constrained by the number of nodes (at least two for fault tolerance in the monitoring system) that provide the accuracy within a bound, to be defined later.

In the following subsections, we present the details of the two stages of the objective function. In the first subsection, we present the details of the two stages of the objective function: the first stage; optimisation of the performance measure-i.e. the accuracy of leakage detection and localisation. In the second subsection, we present the second stage of the objective function aimed at optimising the overall energy consumption.

5.2.1 The First Stage: Accuracy Optimisation

We formalise our decision process using MDP. An MDP can be used as a model for formalising decision processes in a stochastic environment using a set of variables

such as $\langle S, A, p, r \rangle$ [140]. In this tuple, S represents the sets of states that the players can be in the environment. A is the set of available actions taken to transition from one state to another. p represents the probability transition function, while r denotes the reward or utility function. Also important to consider is the discount factor β for future rewards, as will be explained later. Given these sets of variables, the objective is to find an optimal policy that maximises the expected discounted cumulative reward. This can be achieved using the well-known Bellman optimality equation. The Bellman optimality equation enables the measurement of the goodness of a state -i.e. the maximum obtainable reward in a state through a state value function $V(s)$.

In general, when an arbitrary policy is followed, the Bellman expectation equation is defined as follows.

$$\forall s : V^\pi(s) = \sum_{s'} p(s'|\pi(s), s) [(r(s, \pi(s)) + \beta V^\pi(s'))] \quad (5.3)$$

where $V^\pi(s)$ is the value in state s based on policy π , $p(s'|\pi(s), s)$ is the probability of going from state s to s' when action following a policy, $r(s, a)$ equals the immediate reward in state s , β is the discount factor for future rewards and $V^\pi(s')$ is the future value.

Nonetheless, to maximise the accumulated reward, an agent should follow an optimal policy. Unlike the general equation that defines the expected reward by following an arbitrary policy, the optimal policy guarantees the maximum reward for the immediate state and all future states with the assumption that each state continues to follow this policy. Hence, the policy that takes action with the maximum reward is defined by the Bellman optimality equation as follows.

$$\forall s : V^*(s) = \max_a \sum_{s'} p(s'|s, a) [(r(s, a) + \beta V^*(s'))] \quad (5.4)$$

The Bellman optimality equation is divided into two parts: the first part represents the immediate reward denoted by $r(s, a)$. The second part means accumulated future rewards, gathered through the iterative part $V^*(s')$. In the second part of the equation, we use the hyperparameter β . The β variable is vital to this part to avoid infinite cycles and enable solutions' eventual convergences. In addition, it emphasises the importance of future rewards, i.e. the higher the value of β , the more important the value of the long-term reward.

Thus, in the first stage of our work, we use Bellman's optimality equation to evaluate different policies and to determine the policies that maximise the reward in terms of accuracy of leakage detection and localisation. As such, We map the environment (NNPC pipeline network) to the equation as follows:

The States S : In our work, a state is defined as the conditions based on which leakage detection and localisation process are applied. Considering the regionalisation defined in the previous section, we define our set of state S as follows: $S = \{start, con, r_0, r_1, r_2\}$. The states *start* and *con* represent the initial states where no failure is present. The latter state represents communication initialisation. This includes the connection definition with predefined neighbours discussed in the previous chapter. States r_0, r_1, r_2 denotes the states during failure. As earlier defined, these states depend on the geographical area's failure rate.

The Actions A : Actions are used to transition between states. In this work, we have several actions that allow this transition. They are represented by the set $A = \{s_t, c, s_a, r_{ds}, m_{ds}, rm_{ds}\}$ representing the start process, connection initialisation, service activation, replication of data and services, migration of data and services and replication and migration of data and services respectively. Services in our work include data preprocessing, leakage detection, leakage localisation, data transfer/ data sharing, data filtering, data aggregation and prioritisation and mid-term data storage. Only the first four service can be done at the sensor node level. At the fog layer using the gateways, all services except long-term historical data storage are implemented.

s_t (start-action): the start-action is a switch-on action symbolising the beginning of the decision process.

c (connection): the connection initialisation action is used to implement the connectivity between nodes as defined in HyDiLLEch in the previous chapter.

s_a (service activation): this action is used to activate services in the nodes. Services are deployed in two ways in this work: service pre-deployment and dynamic service deployment. In the case of service pre-deployment, the service switch action is a mechanism to reduce energy consumption by intermittently activating services as needed. It is, thus, applicable alongside other actions at every point in time. Dynamic service deployment, on the other hand, allows strategic placement of services in sensor nodes, gateways or the cloud.

r_{ds} (data and service replication): this action denotes the replication of data and services as the name implies. This action is taken by the creation of copies of data as they are being produced or services in nearby sensor nodes. Additionally, as discussed in the previous chapter, it incorporates replication to the fog nodes instead of the limitation to the sensor node. The implementation of other services, such as alarms, data prioritisation and filtration, to mention a few, is enabled by the extension of the replication action to fog nodes.

m_{ds} (data and service migration): involves the migration of data and services. While considerably similar to r_{ds} , m_{ds} differs from r_{ds} as it is not only used in the presence of failures, it is also used as a response to specific deployment needs. For example, it can be used to reassign services such as data aggregation and storage or

data prioritisation to other nodes in case of memory exhaustion in the hosting node. In addition, we can also make use of the migration action when there is an increased latency between a particular service and the data it requires. Note that both r_{ds} and m_{ds} have several, but different, communication requirements, as will be discussed later.

rm_{ds} (data and service replication and migration): In areas where the failure rate is exceptionally high, we can utilise a combination of actions as a single action. Therefore we include the rm_{ds} action, which is a combination of both replication and migration actions to be taken as a single action in this case.

The States Transition Function p : The state transition function p represents the probability of moving from one state to another when the environment dynamic is known. In our work, we utilised the information from the data presented in Fig. 5.2 to define this transition. According to this data, some geographical areas will likely transition from one region to another. Thus, we determine how each geographical area moves within the logical regions.

The Reward Function r : The reward function is defined as $r \in [0, 100]$ per node for detection and localisation accuracy of leakages. The maximum obtainable reward for every action taken is equivalent to the accuracy of leakage localisation. In addition, we consider the number of nodes that falls within the allowance threshold (to be defined later) for an acceptable accuracy level representing the aspect of fault tolerance.

5.2.2 The Second Stage: Optimising the Energy Consumption

The second stage of the work includes the selection of policies that minimise the energy consumption for each region from the results of stage 1. We determine the energy consumption from actions on the environment by representing the interaction of the nodes in the environment as follows:

Let us define a set of nodes $N = \{n_1, n_2, \dots, n_u\}$, where u - the total number of nodes (sensors nodes and gateways). Each node $n_i \in N$ produces a set of data used by the services deployed in the system, which are also stored in the nodes. If G is a matrix for the use of data by each service, then $g_{ij} = 1$ represents service in node n_i requiring data produced by node n_j . As such, we propose h_{ij} to equate the number of hops between these nodes.

$$G = \begin{bmatrix} g_{11} & \cdot & \cdot & g_{1u} \\ \cdot & & & \\ \cdot & & & \\ g_{u1} & & & g_{uu} \end{bmatrix} \quad (5.5)$$

$$NE = \begin{bmatrix} ne_{11} & . & . & ne_{1u} \\ . & & & \\ . & & & \\ ne_{u1} & & & ne_{uu} \end{bmatrix} \quad (5.6)$$

Matrix NE is the predefined neighbourhood between the nodes such that $ne_{ij} = 1$ represents a data sharing between nodes and 0 otherwise.

There are several modes of communication between sensor nodes and gateways. Let us represent the communication modes by $l_{ij} \in \{0, 1, 2\}$ where $l_{ij} = 0$ where $l_{ij} = 0$ is communication absence, $l_{ij} = 1$ a connection via LoRa and $l_{ij} = 2$ a connection via 3G/4G communication networks. The communication modes defer in their capacity and influence the possible action between nodes. For example, the service activation is a straightforward action that depends only on the predefined neighbourhood connection, i.e. the implementation of state *con*.

When the replication is considered, the communication between originating and destination nodes must be $l = 1$ for both data and services.

In the case of the migration action, the communication requires $l \in \{1, 2\}$ for data and service, respectively. This is because the migration of services needs higher bandwidth than that of data.

Thus, the energy consumption for each action is calculated using the following the following relation:

$$cost_{ij} = lc * h_{ij} \quad (5.7)$$

where lc is the cost per packet, and per link of the service or data, h_{ij} is the number of hops between the origin and destination nodes.

Note that when services are pre-deployed in all nodes, the energy cost change slightly, i.e. the cost of service migration will not be incorporated. In this case, the service activation action switch s_a will be used in place of the service migration action. Given this communication mode, we can minimise energy consumption by reducing the distance (in the number of hops) between services and the needed data to run the service. Note that *a good placement* of service is when we have it placed at most one hop away from where leakage occurs. Finally, we can define the objective of the second stage as follows:

$$\min_{(\pi^*, V^*)} V_e(\pi^*, V^*) = \min_{(\pi^*, V^*)} \left[\sum_{i=1}^u \sum_{j=1}^u cost_{ij} \cdot \sum_{s'} \phi^\pi(s') \sum_s g_{ij}(s'|\pi(s), s) \cdot V^*(s') \right] \quad (5.8)$$

where $\phi^\pi(s') = p(s'|s, \pi)$ is the steady state defined by the probability of moving to the next state following a policy (π) in the current state, $cost_{ij}$ is the energy consumed defined in Eqn. 5.7, $g_{ij}(s'|\pi(s), s)$ represents if data/service is resident in different nodes when we follow a policy from one state to another.

This equation presents our objective function, i.e. to find and ensure convergence to the optimal reward with minimum energy consumption through the various placement strategies in the network nodes.

5.3 Implementation and Results

The model is implemented using the Gym Library. It is an open-source toolkit that can be used to develop and compare reinforcement learning algorithms. The NNPC environment shown in Fig. 5.1 was simulated using the NS3 network simulator. To realise the communication between Open AI Gym and NS3, we build on the work conducted by Gawlowicz *et al.* [142] called ns3-gym. This work allows seamless communication between the OpenAI Gym framework and the NS3 network simulator. The interaction is realised using an instantiated gateway in the NS3 environment and a proxy in the Gym environment. Our implementation of the interaction between the simulated pipeline network and the OpenAI Gym is depicted using Fig. 5.4. We used multiple agents on the Gym side for comparative analysis. To implement the placement strategies for the regions while carrying out the reinforcement learning using the algorithm Alg. 3, each agent interacts with a gateway through the dedicated proxy.

The information exchanged between an agent and a gateway in each episode are: (i) observation space, (ii) the action space, (iii) the reward and (iv) the game over conditions. In our case, we defined them as follows:

The observation space: Failures, i.e. leakages, network or communication failures.

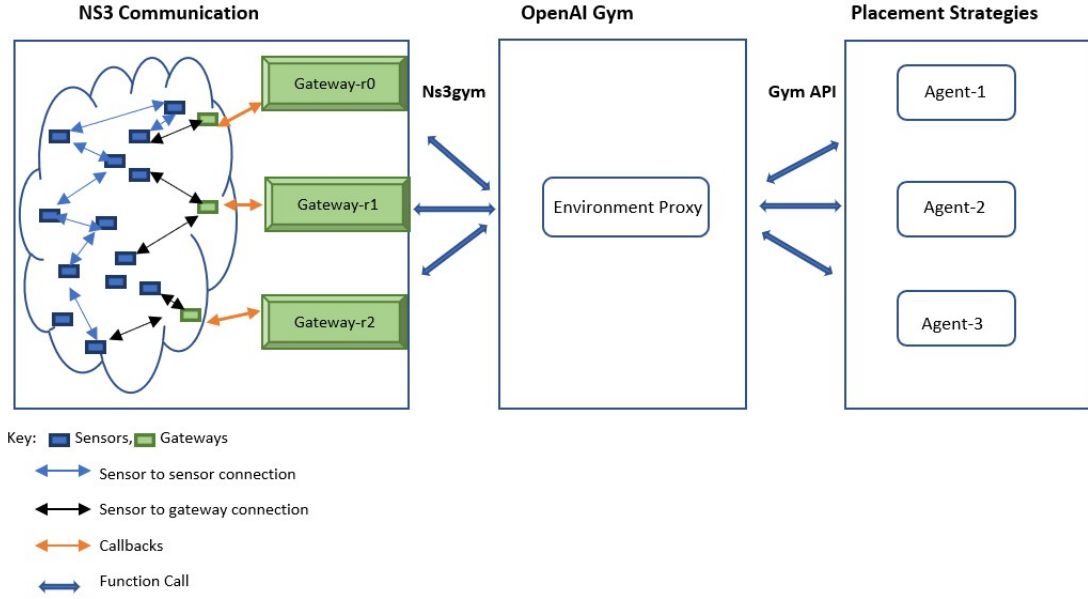


Figure 5.4: Simulated pipeline network and OpenAI Gym Interaction

The action space: Actions include *start*, *connection*, *service activation*, *data and service replication* and *data and service migration*

Reward: The reward is based on the localisation accuracy of the leakage and the number of nodes that localises the leakages.

Game over: A game is considered over at the end of the episodes.

We present the simulation parameters in Table 5.1.

Timestep	1000
Areas	5
Region	3
Proxy	1
Alpha	1
Beta	0.99
Number of episodes	1000
Reward per node	[0,100]
Error Type	Rate error model, list error model
epsilon	[0, 1]

Table 5.1: Simulation Parameters

The Q-learning algorithm is a well-known reinforcement learning algorithm that guarantees the convergence of an MDP using greedy policies to return the maximum rewards for the trajectories of the states. In our work, however, we use the *epsilon*-greedy Q-learning algorithm to obtain the optimal value function for the three regions. This version of the Q-learning algorithm shown in Alg. 3 and extracted from [141]

Algorithm 3 ϵ -greedy Q-learning

```
1: Reset Environment
2:  $\epsilon \in (0, 1]$ ,  $\alpha = 1$ ,  $\beta = 0.99$ 
3: Initialise  $Q(s,a)$ ,  $E(s,a)$  for all  $s \in S$ ,  $a \in A(s)$  arbitrarily except  $Q(\text{terminal},.)=0$ 

4: for each timestep in each episode do
5:   Initialise  $s$ 
6:   for each region do
7:     Choose  $a$  from  $S$  using policy ( $\epsilon$ -greedy) derived from  $Q$ 
8:     Perform action  $a$ 
9:     Observe  $r$ ,  $s'$ 
10:     $Q(s, a) \leftarrow Q(s, a) + \alpha[R + \beta \max_a Q(s', a) - Q(s, a)]$ 
11:     $s \leftarrow s'$ 
12:    Get corresponding energy consumption using 5.7 {Update  $E$ }
13:   end for
14:   Until  $s$  is terminal
15: end for
16: Reset Environment
17: For each region, choose the policy with the least energy consumption and optimal reward
```

ensure a balance between exploration and exploitation of states. We use this algorithm because, given the nature of our problem, whereas an optimal policy guarantees convergence to an optimal value, the rate of incidents across the regions differs significantly. Thus, for each region, we solve the MDP such that the unique policy returned from the solution is that which not only satisfies optimality for that region but also minimises the global energy consumption.

Therefore, we take our environment, the NNPC environment and define the region of each of its areas in Alg. 3. This is done by setting the corresponding failure of the area using rate and list error types for the regions. Using a matrix $-Q-$, we store the state-action values and keep track of the energy consumption using the corresponding energy consumed with a matrix E . At the end of the episodes, we select the optimal policy for each region.

We discuss the simulations and results in the following subsections:

5.3.1 Accuracy in Detection and Fault Tolerance

We evaluate the *fault-tolerance* and *accuracy* of the LDMS (from the previous chapter) for all regions in this subsection. Given the context of our, these two are jointly considered the *obtainable reward* for the MDP. We examine both metrics using two types of communication failures in NS3, i.e. the *rate error type* and *list error type*. While both methods can introduce failure to the system, they differ in how they work. The rate error type introduces failure via random packet drops, delays and out-of-order packet delivery. On the other hand, the list error type involves user-selected packet drops. In our case, we used uniformly selected packets across the list of packets shared among the sensor and gateway nodes. To examine performance, we examine the effects of these failures across randomly selected leakage points. Obtained results are shown in Figures 5.5, 5.6, 5.7, 5.8, 5.9 and 5.10. Note that for each error rate, the evaluation is limited to the range of 0% to 20%. Outside this threshold, detection or localisation becomes impossible. Additionally, each metric of evaluation is presented in two figures (all possible outcomes and controlled outcomes). As the name implies, the former represents the outcomes recorded from all tests across all ranges. The latter, on the hand, represents the outcome of High Performing Nodes (HPN). The HPN are nodes that produce an accuracy of detection greater than 90%. We set the minimum level of accuracy to 90% following the high level of performance required by the oil and gas operators [48]. This set threshold also helps standardise the benchmark based on which performance is measured.

Fault tolerance

One of our LDMS's strengths is removing Single Points of Failure associated with centralised systems. Thus to maintain this property, we need to ensure that there are at least two HPNs across all regions. Figures 5.5 and 5.6 shows the number of nodes that can detect and localise leakages.

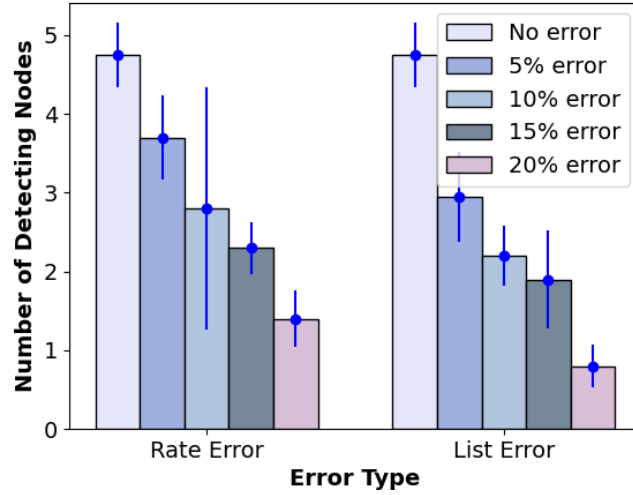


Figure 5.5: Total Number of Nodes Detecting Leakage

While Fig. 5.5 represents the total number of nodes that can detect and localise leakage without putting into consideration the level of accuracy, Fig. 5.6 captures only the HPN, i.e. with accuracy over 90%. In Fig. 5.6, we observe that areas with 5% failure rate for both rate and list error types maintain the required number (at least two) of nodes for fault tolerance, mainly differing in the variance. Additionally, we also observe that this variance is from nodes located at the extremities of the pipelines in the area(s) in consideration. This problem can be addressed by increasing the node density in such locations. Finally, we find that other areas with a failure rate above 5% do not meet this requirement.

Accuracy

In addition to the fault tolerance level, we also examine how the accuracy level changes based on the failure rate. Results are presented in Figures 5.7 and 5.8.

While the fault tolerance decreases as the failure rate increases, the accuracy level presents a differing result. Besides, the error types differ as well. For example, in Fig. 5.7, we notice that the list error type provides a higher accuracy level than

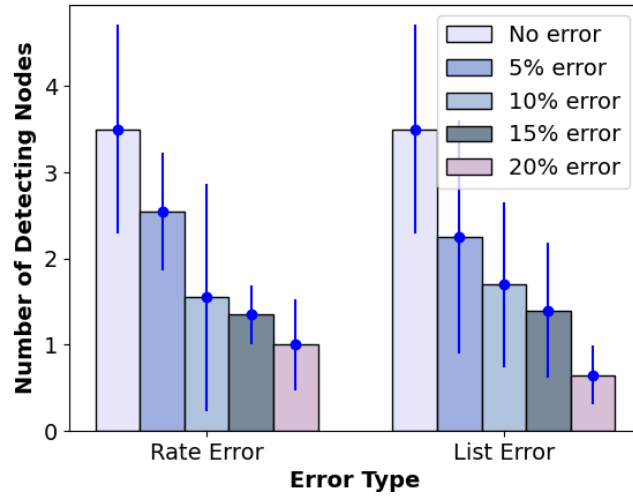


Figure 5.6: Total Number of HPN



Figure 5.7: Accuracy in Leakage Detection from all nodes



Figure 5.8: Accuracy in Leakage Detection in HPN

the rate error type, with both showing similar trends for the failure rate. Although, when we consider results from HPN shown in Fig. 5.8, the rate error type presents much variance in changes in the accuracy of localisation compared to the list error type. Thus, we can make the inference that while the lower error rates favour a higher number of nodes in leakage detection, it does not have such an impact on the accuracy of localisation.

Rewards

Therefore, we base the definition of the reward function on the minimal requirements of both fault-tolerance and accuracy levels. Thus, the total reward is the sum of the accuracy level obtained by each detecting node in the region. We also represent the reward function in Figures 5.9 and 5.10 as all outcomes and HPN outcomes. Both cases show a similar pattern of decrement in reward as the failure increases.

Finally, we present the expected reward across the regions in Fig. 5.10 following the benchmark discussed in subsection 5.3.1. For each region, we expect to have two HPNs at all times for all failure rates, thus putting the target at the level shown in the figure.

5.3.2 Optimised Accuracy and Energy Consumption

We implemented several algorithms (learning-based, heuristic and pessimistic), thus, in this subsection, we discuss the rewards obtained based on these implementations.

For the learning-based algorithms, we used a uniform random policy and our



Figure 5.9: Total Obtainable Reward across the Error Rates



Figure 5.10: Total Obtainable Rewards with HPN

proposed regionalised ϵ -greedy learning (R-MDP). While the uniform random policy has a uniform distribution over the space from which actions are taken, R-MDP finds the balance between exploitative and exploratory actions. This is possible using the ϵ -greedy Q-learning algorithm, which allows the agent to exploit what has been learnt and explore actions with yet-to-be-known outcomes. This strategy of balancing between exploitation and exploration avoids local optimum.

We provided the baseline for comparative analysis by implementing a heuristic approach based on Weighted Set Cover (WSC) service placement from the work [143]. The aim is to maximise detection accuracy with a minimal cost following the context of our work. We also included a pessimistic approach based on the worst-case scenario as a performance measure. Each algorithm is evaluated based on the value function and the energy consumption.

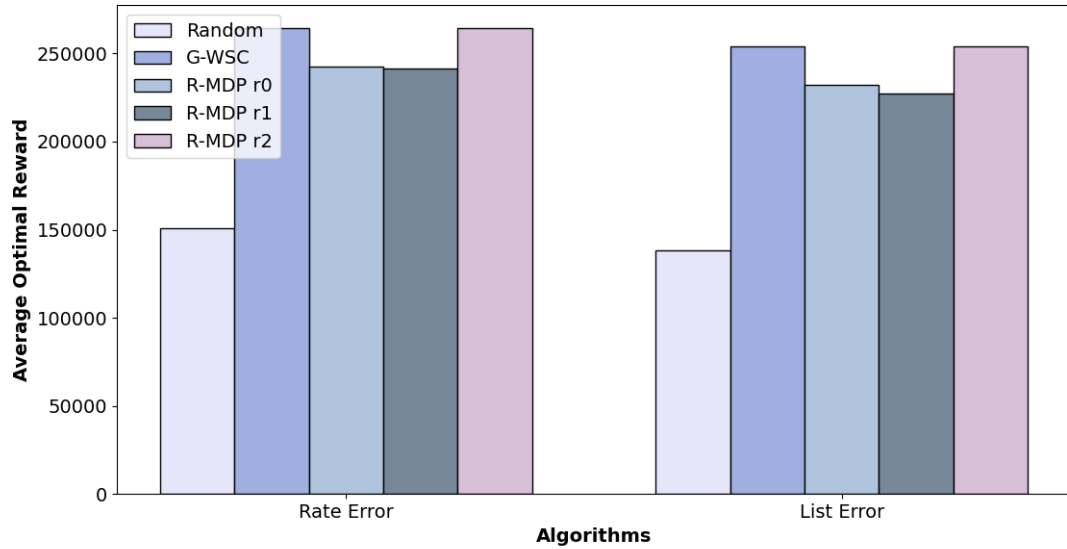


Figure 5.11: Total Reward by Algorithm

We present the optimal reward for the implemented algorithms in Fig.5.11. The random policy expectedly has the lowest total reward. The other algorithms, however, show a much better performance. The R-MDP for each region shows a similar reward across the three regions differing slightly. The WSC-based algorithm, however, shows an increased reward of about a 6% rise in the obtained total value for the R-MDP regions.

While the optimal value functions are similar, obtained energy results show a significant difference in energy consumption due to the exploration effects. We present

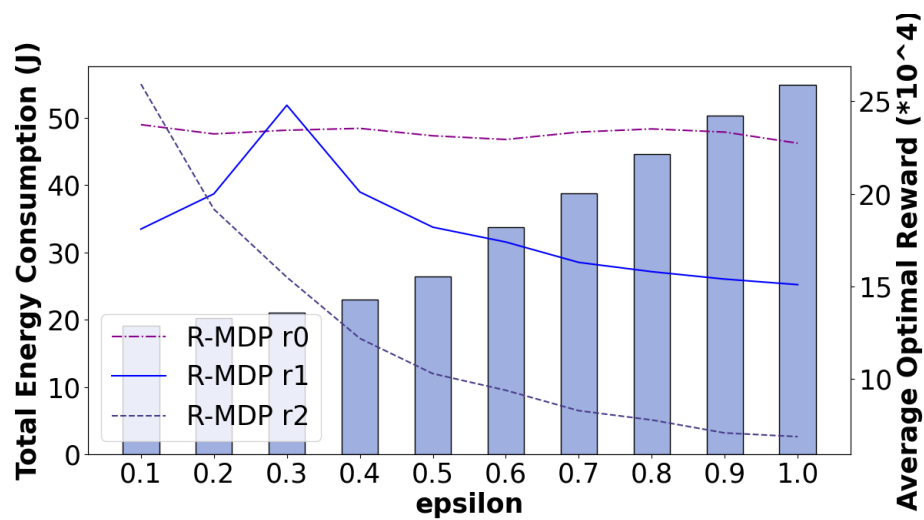


Figure 5.12: Total Energy Consumption by Algorithm

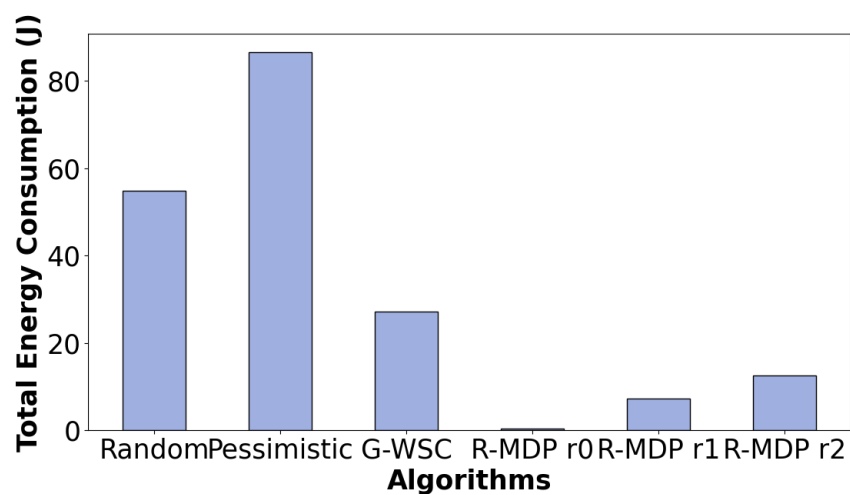


Figure 5.13: Regional Reward vs Energy Consumption

the total energy consumption and the corresponding optimal value in Fig. 5.12. In the figure, the energy consumed is presented as the histogram while the value function is presented in lines. According to the result, we observe lesser energy consumption for average epsilon values between 0 and 0.4 and an increased energy consumption as the epsilon value approaches 1. However, these effects differ considerably for each region, i.e. regions $r0$ and $r2$ have a better performance with near greedy or greedy policies than more explorative policies. Although the reward for region $r0$ increases slightly as the exploration increases, the reward for region $r2$ is inversely proportional to the value of epsilon. In the case of region $r1$, the optimal result is achieved with the epsilon value equaling approximately 0.3. Hence, we can conclude that partitioning the areas of the environment into several regions allows an optimal solution with a minimised energy consumption.

Using Fig. 5.13, we present the energy consumption by the algorithm. In this presentation, we analyse the total energy consumed by the different algorithms. Results show that the pessimistic algorithm expectantly has the highest energy consumption. In comparison to the total energy consumed by the R-MDP, this approach consumes approximately 77% more energy. However, we observe a 26% reduction in energy consumption when we compare it with the globalised heuristic approach (WSC). While this is much lesser than the reduction obtained compared to the pessimistic approach, the decrement in this case still poses a significant improvement.

5.4 Conclusion and Discussion

This chapter presented our work on fault-tolerant and energy-efficient data and service management modelled as an MDP. To model the MDP, we used the NNPC pipeline network and the historical failure pattern over a five-year period. Leakage detection and localisation are done using LDMS from the previous chapter. In this chapter, we took into consideration the data and service layer as opposed to just the sensor layer. We showed that energy consumption could be made efficient when we consider the dynamics of the environment. In the previous chapter, we used a generic scenario for the detection and localisation of leakages in a pipeline of twenty kilometres. We did this without considering any failure. In this chapter, however, we consider different types of failures, i.e. the rate and list error rates failures in a more dynamic environment. We observe that considering this aids the efficiency of global energy consumption.

Chapter 6

Conclusion and Future Works

This chapter presents a summary of our research and obtained results in 6.1. We follow this with a discussion on the possibilities of future works in section 6.2.

6.1 Summary

This thesis presents our work on a resilient IoT-based solution for addressing failures in the oil and gas industry. In our, we focus on the midstream sector of Nigeria's oil and gas industry, mainly the failures in pipeline transportation of crude oil. We proposed an IoT-based solution and considered a multi-layer fault tolerance published in [118]. To implement our contribution, we first assessed the design and specification aspects of the system. We proposed a three-layer hierarchical architecture consisting of the WSN layer, the fog layer and the cloud layer.

At the WSN Layer, we implemented our first contribution published in [130]. This included the design and simulation of a hybrid and distributed leakage detection and localisation technique-HyDiLLEch. HyDiLLEch combines three existing LDTs (PPA, GM and NPWM), each technique with its advantages and disadvantages. For instance, while the PPA can be used to detect leakages only, GM can be used to detect and localise leakages. However, the accuracy of localisation depends on the sensors. NPWM, on the other hand, requires a high sampling rate for accurate localisation. Thus, HyDiLLEch aimed at taking advantage of each detection technique's strengths while minimising its weaknesses. With HyDiLLEch, we are also able to detect multi-sized leakage in a manner that eliminates SPOF from the detection system using a unique node placement strategy based on crude oil propagation. Our results showed that we could detect leakages with a high accuracy level without high energy consumption cost compared to some classical methods, such as the NWPM. However, compared to the other method GM, the most significant improvement is in the area of fault tolerance, with a four to six increment to the number of nodes detecting leakages.

At the fog layer, we considered efficient data and service placement for a fault-tolerant and energy-efficient LDMS published in [136]. This problem was modelled as an MDP based on the historical failure data of the NNPC pipeline network. We used various placement strategies, such as replication and migration of data and/or service between the sensor and fog nodes. With five distinct areas, the NNPC pipeline network presented diverse failure rates. Thus in our model, we regionalised (r0, r1, r2) the

MDP to provide a similar level of performance across the different areas of the NNPC pipeline network using an ϵ -greedy algorithm for balancing the exploration and exploitation actions. With this regionalisation, we find that each region reaches its optimal value function following a different strategy. For regions r0 and r2, the policy adopted is near greedy policies. Although both behave differently across the epsilon-value spectrum, the optimal value for both regions is closer to the minimum value of zero. The most dynamic region of all is the r1 region. In this region, we find that the optimal value function is obtained with an epsilon value of approximately 0.3. These different regional strategies allow the optimisation of the value function while minimising energy consumption.

The cloud layer is used to store historical data and host infrequently used services such as alarm services. More heavy tasks outside the replication, migration, detection services, and localisation services. Unlike the first two layers, connection to the cloud requires a more robust connection through the backbone network, such as the 3G/4G and possibly 5G networks in the future.

Overall, our work addressed some issues across the layers of an IoT-based LDMS. We implemented an LDMS that is robust to SPOF and communication failure at the network layers, such as packet drops, out-of-order delivery, and communication delays. We also considered the scalability and energy efficiency of the LDMS by addressing the connectivity of nodes and data sharing amongst geographically close sensors as well as policy-based data and service management.

6.2 Future Works

Our work addressed several aspects of efficiency in an IoT-based monitoring system for crude oil pipelines in Nigeria. Still, a multitude of opportunities exists to further the scientific research for such systems.

To evaluate the system's fault tolerance, we considered failures at the network layer. These failures include rate and list error type, which concerns packet drops, delay, and out-of-order delivery. Other communication failures, such as node failures, and propagation losses at the physical level, could be considered for further evaluation and possible enhancement. Additionally, more simulations to check the conformation of obtained results for the physical layer and the overhead of LoRaWAN services can be conducted using the 2021 release of the LoRaWAN module in NS3-an ongoing work of Magrin *et al.* and Capuzzo *et al.* [144,145]

Other options to enhance the system performance include the node placement strategy. For instance, we have used a homogeneous node placement across all the pipeline networks. This resulted in suboptimal detection at some extremities as discussed in chapter 5. Heterogeneous node placement based on location, in addition to

the fluid propagation properties, can resolve such problems.

In addition, we proposed an energy-efficient regionalised single-player game for data and service management due to the high diversity in the historical data of the NNPC pipeline networks over the geographical regions. However, various factors contributed to such disparity. As such, unsupervised learning-a machine learning technique- can be applied as a pre-processing step to determine the pattern or structure of the failure data according to the causes. This step can provide a more refined categorisation for causal modelling of the problem for an enhanced solution.

Also, such diversity as seen in the NNPC pipeline network historical data is absent in other countries, such as the United States, recording high incident rates in its pipeline networks. Besides, the analysis of pipeline incidents in [146] shows a more homogeneous pattern across different factors in various countries. In such situations, a two-player game would be a more practical solution for a highly vandalised environment using a competition-based model such as the Stackelberg model utilised in [115]. A zero-sum game applying a score-based approach would also provide a stable solution on the assumption that each player is rational [140]. Yet, this assumption cannot be guaranteed especially as it concerns third-party interference. Other options include adopting the Schrödinger-based model as in the work [147]. In this work, it is assumed that each player has only two states, where each state represents a different solution to the game. Playing conditions also include the time duration and outside factor influences. However, such refined approaches can only be applied to specific cases. As the situation and failure rates and types vary significantly across oil-producing countries, a better approach could be the adoption of a hybrid system using n players. In this case, n could be determined based on the historical data considered, thereby providing a generic solution to data and service management in crude oil pipelines.

Furtherance to that, the design of a specialised multi-sensory device that includes GPS, pressure sensor, and speed sensor is required to carry out experimentation of our work. While the algorithm itself has been proven to be energy efficient, other opportunities exist to improve the energy efficiency in the design phase of such a sensing device. In [148–150], Bonvoisin *et al.* proposed a framework for the analysis of the environmental impact of WSN and ICT solutions throughout their lifecycle. These include sensor devices, gateways, interaction models, and optimisation at the various levels of the system. Additionally, Achachlouet, in his work [151], enumerate the importance of the lifecycle approach to the design of ICT solutions for environmental sustainability from its direct and rebound effects.

Hence, a lifecycle-based approach to creating such a sensor could tremendously reduce energy consumption and consequently the environmental impact.

Options to consider range from the source of the electrical components as well as the communication aspect. For example, the communication aspect of the device can be based on improved LoRaWAN classed devices. LoRaWAN devices act as transceivers communicating with each other or the gateway on well-defined transmission and receive windows for class A, B or C devices. HyDiLLECh also works with defined intervals setting precedence for research on suitable device classes during the design phase.

According to the United States energy information administration [152], crude oil currently accounts for one-third of global oil consumption. While this figure is forecasted to be on the rise, alternative energy sources such as liquefied natural gas (LNG), biofuels and renewable energy are also increasingly considered. LNG, in particular, is a fast-rising alternative energy source due to its minimal contribution to greenhouse gases and elevated combustion efficiency [153]. Yet, like crude oil, LNG is transported via pipelines [154, 155], which are also susceptible to failures resulting in environmental hazards [156]. Hence, as a long-term opportunity ensuring a fault-tolerant LDMS for LNG is paramount. Such a system can be built using our LDMS as a model especially the architectural design and communication aspects.

Furthermore, water pipeline monitoring is an essential aspect of the development of smart cities [157, 158]. Current approaches to its monitoring are based on centralised WSN [58, 159], making them susceptible to SPOF. As a result, a distributed approach to pipeline monitoring, such as HyDiLLECh, could be applied.

Another possibility to consider is the design and development of an IoT-based ecosystem for monitoring the processes in the three sectors of the OGI. As the sectors are codependent, a holistic approach to monitoring all the sectors will significantly improve the overall system. For instance, oil and gas transportation originates from the wellhead, storage facilities or offshore drilling source - vital parts of the upstream sector in the OGI. The transportation aspect of the product begins from such sources using the gathering lines. Hence, monitoring such components as the oil well is essential for the overall supply chain [160]. Likewise, the operational and supply chain activities across the sectors can be enhanced through efficient data analysis from an IoT-based system [161, 162]. For integrated and preventive asset integrity maintenance and management across the sectors, AI-based and machine learning techniques such as Bayesian network can be used. Such possibilities can only be enabled through efficient and long term data collection using IoT-based monitoring systems for incidents in the three sectors.

Bibliography

- [1] B. WETZEL, “Oil In Motion: How Crude Oil Transportation Works,” Breakthrough, Tech. Rep., 2019.
- [2] PHMSA, “Pipeline failure causes,” U.S. Department of Transportation, Tech. Rep., 2019.
- [3] UNEP, “Environmental assessment of ogoniland,” United Nation Environment Programme, Tech. Rep., 2012.
- [4] A. Slaughter, A. Mittal, and V. Bansal, “Bringing the digital revolution to midstream oil and gas,” Tech. Rep., 2018.
- [5] R. I.-T. Y.2060, “Overview of the internet of things,” International Telecommunication Union, Tech. Rep., 2012.
- [6] M. S. Mahmoud and A. A. H. Mohamad, “A study of efficient power consumption wireless communication techniques/ modules for internet of things (IoT) applications,” *Advances in Internet of Things*, vol. 06, no. 02, pp. 19–29, 2016.
- [7] A. Ambituuni, E. Ochieng, and J. M. Amezaga, “Optimizing the integrity of safety critical petroleum assets: A project conceptualization approach,” *IEEE Transactions on Engineering Management*, vol. 66, no. 2, pp. 208–223, may 2019.
- [8] NNPC, “2019 annual statistical bulletin,” Nigerian National Petroleum Corporation, Tech. Rep., 2019.
- [9] R. Amponsah and F. K. Opei, “Ghana’s downstream petroleum sector: An assessment of key supply chain challenges and prospects for growth,” *International Journal of Petroleum and Oil Exploration Research*, vol. 1, no. 1, pp. 001–007, 2014.
- [10] D. Koroteev and Z. Tekic, “Artificial intelligence in oil and gas upstream: Trends, challenges, and scenarios for the future,” *Energy and AI*, vol. 3, p. 100041, 2021.
- [11] A. Sircar, K. Yadav, K. Rayavarapu, N. Bist, and H. Oza, “Application of machine learning and artificial intelligence in oil and gas industry,” *Petroleum Research*, 2021.
- [12] A. Gezdur and J. Bhattacharjya, “Digitization in the oil and gas industry: Challenges and opportunities for supply chain partners,” in *Working Conference on Virtual Enterprises*. Springer, 2017, pp. 97–103.

- [13] C. Lima, S. Relvas, and A. P. F. Barbosa-Póvoa, "Downstream oil supply chain management: A critical review and future directions," *Computers & Chemical Engineering*, vol. 92, pp. 78–92, 2016.
- [14] S. Lisitsa, A. Levina, and A. Lepekhin, "Supply-chain management in the oil industry," in *E3S Web of Conferences*, vol. 110. EDP Sciences, 2019, p. 02061.
- [15] SCL, "Crude oil transportation – pipes, rail, trucks and ships," Southern Counties Lubricants, LLC, Tech. Rep., 2022.
- [16] GLCDGL, "Crude oil transport: Risks and impacts," *Great Lakes Commission*, 2015.
- [17] T. Pootakham and A. Kumar, "A comparison of pipeline versus truck transport of bio-oil," *Bioresource technology*, vol. 101, no. 1, pp. 414–421, 2010.
- [18] B. Blair, J. Hughes, W. Allshouse, L. McKenzie, and J. Adgate, "Truck and multivehicle truck accidents with injuries near colorado oil and gas operations," *International Journal of Environmental Research and Public Health*, vol. 15, no. 9, p. 1861, aug 2018.
- [19] A. Ambituuni, P. Hopkins, J. Amezaga, D. Werner, and J. Wood, *Risk Assessment Of A Petroleum Product Pipeline In Nigeria: The Realities Of Managing Problems Of Theft/sabotage*, ser. WIT Transactions on The Built Environment. WIT Press, 2015, vol. VI, pp. 49–50.
- [20] K. D. Retzer, R. D. Hill, and S. G. Pratt, "Motor vehicle fatalities among oil and gas extraction workers," *Accident Analysis & Prevention*, vol. 51, pp. 168–174, mar 2013.
- [21] S. Ghaleh, M. Omidvari, P. Nassiri, M. Momeni, and S. M. M. Lavasani, "Pattern of safety risk assessment in road fleet transportation of hazardous materials (oil materials)," *Safety Science*, vol. 116, pp. 1–12, jul 2019.
- [22] A. Ambituuni, J. Amezaga, and E. Emeseh, "Analysis of safety and environmental regulations for downstream petroleum industry operations in nigeria: Problems and prospects," *Environmental Development*, vol. Volume 9, p. Pages 43–60, 01 2014.
- [23] C. C. Oriji, "Environmental crude oil pollution and its rising levels of socio-economic/health vulnerabilities among the rural people in rivers state, nigeria," *International Journal of Humanities, Social Sciences and Education*, vol. 5, no. 8, 2018.

- [24] K. Habich, “Understanding and eliminating atmospheric benzene pollution in pasadena, tx,” 2022.
- [25] “Benzene and cancer risk,” American Cancer Society, Tech. Rep., 2016.
- [26] S. Rashid, U. Akram, S. Qaisar, S. A. Khan, and E. Felemban, “Wireless sensor network for distributed event detection based on machine learning,” in *2014 IEEE International Conference on Internet of Things (iThings), and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCoM)*, Sep. 2014, pp. 540–545.
- [27] A. Slaughter, G. Bean, and A. Mittal, “Connected barrels: Transforming oil and gas strategies with the Internet of Things,” Deloitte Center for Energy Solutions, Tech. Rep., 2015.
- [28] T. R. Sheltami, A. Bala, and E. M. Shakshuki, “Wireless sensor networks for leak detection in pipelines: a survey,” *Journal of Ambient Intelligence and Humanised Computing*, vol. 7, pp. 347–356, Mar. 2016.
- [29] SPDCN, “Security, theft, sabotage and spills,” Shell Petroleum Development Company of Nigeria Limited, Tech. Rep., 2017.
- [30] W. Z. Khan, M. Y. Aalsalem, M. K. Khan, M. S. Hossain, and M. Atiquzzaman, “A reliable Internet of Things based architecture for oil and gas industry,” in *19th International Conference on Advanced Communication Technology (ICACT)*, Feb 2017, pp. 705–710.
- [31] S. Shoja and A. Jalali, “A study of the Internet of Things in the oil and gas industry,” in *4th International Conference on Knowledge-Based Engineering and Innovation (KBEI)*, Dec 2017, pp. 0230–0236.
- [32] M. Y. Aalsalem, W. Z. Khan, W. Gharibi, M. K. Khan, and Q. Arshad, “Wireless sensor networks in oil and gas industry: Recent advances, taxonomy, requirements, and open challenges,” *Journal of Network and Computer Applications*, vol. 113, pp. 87–97, 2018.
- [33] R. Nawaratne, D. Alahakoon, D. D. Silva, P. Chhetri, and N. Chilamkurti, “Self-evolving intelligent algorithms for facilitating data interoperability in IoT environments,” *Future Generation Computer Systems*, vol. 86, pp. 421–432, Sep. 2018.
- [34] R. Teja, “What is a sensor? different types of sensors and their applications,” Electronics Hub, Tech. Rep., 2021.

- [35] M. Henrie, P. Carpenter, and R. E. Nicholas, "External and intermittent leak detection system types," in *Pipeline Leak Detection Handbook*. Elsevier, 2016, pp. 137–169.
- [36] Lourenco and T. C, "Verification procedures of ultrasonic flow meters for natural gas applications; procedimento de verificacao de medidores ultra-sonicos para gas natural," Dec. 2009.
- [37] Y. Bai and Q. Bai, "Leak detection systems," in *Subsea Pipeline Integrity and Risk Management*. Elsevier, 2014, pp. 125–143.
- [38] Z. M. Hafizi and E. Vorathin, "A temperature-compensated fbg pressure sensor for underwater pipeline monitoring," in *2020 IEEE 8th International Conference on Photonics (ICP)*, 2020, pp. 66–67.
- [39] R. Tennyson, T. Coroy, G. Duck, G. Manuelpillai, P. Mulvihill, D. J. Cooper, P. E. Smith, A. Mufti, and S. Jalali, "Fibre optic sensors in civil engineering structures," *Canadian Journal of Civil Engineering*, vol. 27, no. 5, pp. 880–889, 2000.
- [40] G. Bernasconi and G. Giunta, "Acoustic detection and tracking of a pipeline inspection gauge," *Journal of Petroleum Science and Engineering*, vol. 194, p. 107549, nov 2020.
- [41] L. Guan, Y. Gao, H. Liu, W. An, and A. Nouredin, "A review on small-diameter pipeline inspection gauge localization techniques: Problems, methods and challenges," in *2019 International Conference on Communications, Signal Processing, and their Applications (ICCSPA)*, 2019, pp. 1–6.
- [42] D. V.-K. Le, Z. Chen, and R. Rajkumar, "Multi-sensors in-line inspection robot for pipeflaws detection," *IET Science, Measurement & Technology*, 2019.
- [43] W. P. Eaton and J. H. Smith, "Micromachined pressure sensors: review and recent developments," *Smart Materials and Structures*, vol. 6, no. 5, pp. 530–539, oct 1997. [Online]. Available: <https://doi.org/10.1088/0964-1726/6/5/004>
- [44] J. F. Wendt, *Computational fluid dynamics: an introduction*. Springer Science & Business Media, 2008.
- [45] M. Henrie, P. Carpenter, and R. E. Nicholas, "Real-time transient model-based leak detection," in *Pipeline Leak Detection Handbook*. Elsevier, 2016, pp. 57–89.

- [46] “Equations of state,” in *Developments in Petroleum Science*. Elsevier, 1998, pp. 129–166.
- [47] J. Sowinski and M. Dziubinski, “Analysis of the impact of pump system control on pressure gradients during emergency leaks in pipelines,” *E3S Web of Conferences*, vol. 44, p. 00166, 01 2018.
- [48] P. Ostapkowicz, “Leak detection in liquid transmission pipelines using simplified pressure analysis techniques employing a minimum of standard and non-standard measuring devices,” *Engineering Structures*, vol. 113, pp. 194–205, apr 2016.
- [49] H. Lu, T. Iseley, S. Behbahani, and L. Fu, “Leakage detection techniques for oil and gas pipelines: State-of-the-art,” *Tunnelling and Underground Space Technology*, vol. 98, p. 103249, apr 2020.
- [50] C. H. Tian, J. C. Yan, J. Huang, Y. Wang, D.-S. Kim, and T. Yi, “Negative pressure wave based pipeline leak detection: Challenges and algorithms,” in *Proceedings of 2012 IEEE International Conference on Service Operations and Logistics, and Informatics*, 2012, pp. 372–376.
- [51] J. Wan, Y. Yu, Y. Wu, R. Feng, and N. Yu, “Hierarchical Leak Detection and Localization Method in Natural Gas Pipeline Monitoring Sensor Networks,” *Sensors (Basel)*, 2012.
- [52] P. Stouffs and M. Giot, “Pipeline leak detection based on mass balance: Importance of the packing term,” *Journal of Loss Prevention in the Process Industries*, vol. 6, no. 5, pp. 307–312, oct 1993.
- [53] S. Y. Na, D. Shin, J. Y. Kim, S.-J. Baek, and B.-H. Lee, “Pipelines monitoring system using bio-mimetic robots,” *International Journal of Electrical, Electronic and Communication Sciences*, 2009.
- [54] R. Sujatha and P. P. Hariprasad, “Robot based smart water pipeline monitoring system,” *IOP Conference Series: Materials Science and Engineering*, vol. 906, no. 1, p. 012025, aug 2020.
- [55] J. Kim, G. Sharma, N. Boudriga, and S. S. Iyengar, “Spamms: A sensor-based pipeline autonomous monitoring and maintenance system,” in *2010 Second International Conference on COMMunication Systems and NETworks (COM-SNETS 2010)*, Jan 2010, pp. 1–10.

- [56] R. Beushausen, S. Tornow, H. Borchers, K. Murphy, and D. J. Zhang, "Transient leak detection in crude oil pipelines," in *International Pipeline Conference*, 2004.
- [57] A. Santos and M. Younis, "A sensor network for non-intrusive and efficient leak detection in long pipelines," in *2011 IFIP Wireless Days (WD)*, Oct 2011, pp. 1–6.
- [58] F. Karray, A. Garcia-Ortiz, M. W. Jmal, A. M. Obeid, and M. Abid, "EARN-PIPE: A testbed for smart water pipeline monitoring using wireless sensor network," *Procedia Computer Science*, vol. 96, pp. 285–294, 2016.
- [59] A. Mirzaei, A. Bahrampour, M. Taraz, A. Bahrampour, M. Bahrampour, and S. A. Foroushani, "Transient response of buried oil pipelines fiber optic leak detector based on the distributed temperature measurement," *International Journal of Heat and Mass Transfer*, vol. 65, pp. 110–122, oct 2013.
- [60] A. Kostianoy and D. Solovyov, "Operational satellite monitoring systems for marine oil and gas industry," in *Proceedings of 2010 Taiwan Water Industry Conference, Tainan, Taiwan*, 2010, pp. 28–29.
- [61] A. G. Kostianoy, O. Y. Lavrova, M. I. Mityagina, and D. M. Solovyov, "Satellite monitoring of the nord stream gas pipeline construction in the gulf of finland," in *Oil pollution in the Baltic Sea*. Springer, 2012, pp. 221–248.
- [62] A. Kostianoy, E. Bulycheva, A. Semenov, and A. Krainyukov, "Satellite monitoring systems for shipping and offshore oil and gas industry in the baltic sea," *Transport and Telecommunication*, vol. 16, no. 2, p. 117, 2015.
- [63] A. Smith, "Gas pipeline monitoring in europe by satellite sar," in *Remote Sensing for Environmental Monitoring, GIS Applications, and Geology II*, vol. 4886. International Society for Optics and Photonics, 2003, pp. 257–267.
- [64] E. Dedikov, N. Khrenov, V. Lukyaschenko, R. Salikhov, A. Ponomarev, M. Ovchinnikov, and I. Rodionov, "The russian small satellite for hyperspectral monitoring of gas pipelines," in *Digest of 3rd Int. Symp. Int. Acad. of Astronautics*. Citeseer, 2001, pp. 235–239.
- [65] A. El Hakim, "Internet of things (iot) system architecture and technologies," *White Paper*, vol. 10, 2018.
- [66] H. Rahimi, A. Zibaeenejad, and A. A. Safavi, "A novel iot architecture based on 5g-iot and next generation technologies," in *2018 IEEE 9th Annual Information*

Technology, Electronics and Mobile Communication Conference (IEMCON), 2018, pp. 81–88.

- [67] L. S. Perelman, W. Abbas, X. Koutsoukos, and S. Amin, “Sensor placement for fault location identification in water networks: A minimum test approach,” *Automatica*, vol. 72, pp. 166–176, Oct. 2016.
- [68] L. Sela and S. Amin, “Robust sensor placement for pipeline monitoring: Mixed integer and greedy optimization,” *Advanced Engineering Informatics*, vol. 36, pp. 55–63, Apr. 2018.
- [69] J. Berry, W. E. Hart, C. A. Phillips, J. G. Uber, and J.-P. Watson, “Sensor placement in municipal water networks with temporal integer programming models,” *Journal of Water Resources Planning and Management*, vol. 132, no. 4, pp. 218–224, jul 2006.
- [70] A. Krause, J. Leskovec, C. Guestrin, J. VanBriesen, and C. Faloutsos, “Efficient sensor placement optimization for securing large water distribution networks,” *Journal of Water Resources Planning and Management*, vol. 134, no. 6, pp. 516–526, nov 2008.
- [71] R. Sarrate, F. Nejari, and A. Rosich, “Sensor placement for fault diagnosis performance maximization in distribution networks,” in *2012 20th Mediterranean Conference on Control Automation (MED)*, 2012, pp. 110–115.
- [72] A. Boubrima, F. Matigot, W. Bechkit, H. Rivano, and A. Ruas, “Optimal deployment of wireless sensor networks for air pollution monitoring,” in *24th International Conference on Computer Communication and Networks (ICCCN)*, Aug 2015, pp. 1–7.
- [73] Y. Guo, F. Kong, D. Zhu, A. Saman Tosun, and Q. Deng, “Sensor placement for lifetime maximization in monitoring oil pipelines,” *Proceedings of the 1st ACM/IEEE International Conference on Cyber-Physical Systems, ICCPS '10*, 05 2010.
- [74] O. E. Elnaggar, R. A. Ramadan, and M. B. Fayek, “WSN in monitoring oil pipelines using ACO and GA,” *Procedia Computer Science*, vol. 52, pp. 1198–1205, 2015.
- [75] A. Albaser and U. Baroudi, “Cluster-based node placement approach for linear pipeline monitoring,” *IEEE Access*, vol. 7, pp. 92 388–92 397, 2019.

- [76] R. Li, M. Wenting, N. Huang, and R. Kang, "Deployment-based lifetime optimization for linear wireless sensor networks considering both retransmission and discrete power control," *PLoS ONE*, vol. 12, 11 2017.
- [77] K. Yelmarthi, A. Abdelgawad, and A. Khattab, "An architectural framework for low-power IoT applications," in *2016 28th International Conference on Microelectronics (ICM)*, Dec 2016, pp. 373–376.
- [78] A. M. Sadeghioon, N. Metje, D. Chapman, and C. Anthony, "Water pipeline failure detection using distributed relative pressure and temperature measurements and anomaly detection algorithms," *Urban Water Journal*, vol. 15, no. 4, pp. 287–295, 2018. [Online]. Available: <https://doi.org/10.1080/1573062X.2018.1424213>
- [79] A. M. Sadeghioon, N. Metje, D. N. Chapman, and C. J. Anthony, "Smartpipes: Smart wireless sensor networks for leak detection in water pipelines," *Journal of Sensor and Actuator Networks*, 2014.
- [80] H. Saeed, S. Ali, S. Rashid, S. Qaisar, and E. Felemban, "Reliable monitoring of oil and gas pipelines using wireless sensor network (wsn) — remong," in *2014 9th International Conference on System of Systems Engineering (SOSE)*, June 2014, pp. 230–235.
- [81] K. Yunana, S. O. Adewale, A. A. Alfa, and M. Sanjay, "An exploratory study of techniques for monitoring oil pipeline vandalism," *Covenant Journal of Engineering Technology (CJET) Vol. 1, No. 1, June 2017*, vol. 1, no. 1, Jun. 2017.
- [82] A. C. Azubogu, V. E. Idigo, S. U. Nnebe, O. S. Oguejiofor, and S. E., "Wireless sensor networks for long distance pipeline monitoring," *International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering*, vol. 7, no. 3, pp. 285 – 289, 2013.
- [83] N. F. Henry and O. N. Henry, "Wireless sensor networks based pipeline vandalism and oil spillage monitoring and detection: Main benefits for nigeria oil and gas sectors," *The SIJ Transactions on Computer Science Engineering & its Applications (CSEA)*, vol. 3, no. 1, Jan. 2015.
- [84] F. C. Obodoeze, H. C. Inyiama, and V. Idigo, "Wireless sensor network in niger delta oil and gas field monitoring: The security challenges and counter-measures," *International Journal of Distributed and Parallel Systems (IJDPS)*, vol. 3, no. 6, Nov. 2012.

- [85] H. Jamali-Rad, X. Campman, I. MacKay, W. Walk, M. Beker, J. van den Brand, H. Jan Bulten, and V. van Beveren, “IoT-based wireless seismic quality control,” *The Leading Edge*, vol. 37, pp. 214–221, 03 2018.
- [86] H. Rudeš, I. N. Kosović, T. Perković, and M. Čagalj, “Towards reliable IoT: Testing lora communication,” in *2018 26th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, Sep. 2018, pp. 1–3.
- [87] D. Ratasich, F. Khalid, F. Geissler, R. Grosu, M. Shafique, and E. Bartocci, “A roadmap toward the resilient Internet of Things for cyber-physical systems,” *IEEE Access*, vol. 7, pp. 13 260–13 283, 2019.
- [88] T. Rebbeck, “M2m and internet of things (IoT). opportunités for telecom operators,” *Analysys Mason*, Jan. 2016.
- [89] A. Celesti, A. Buzachis, A. Galletta, M. Fazio, and M. Villari, “A nosql graph approach to manage IoTaaS in cloud/edge environments,” in *2018 IEEE 6th International Conference on Future Internet of Things and Cloud (FiCloud)*, Aug 2018, pp. 407–412.
- [90] H. FATIMA and K. WASNIK, “Comparison of sql, nosql and new sql databases in light of internet of things – a survey,” *International Journal of Advances in Electronics and Computer Science*, vol. 3, Mar. 2016.
- [91] S. Rautmare and D. M. Bhalerao, “Mysql and nosql database comparison for iot application,” in *2016 IEEE International Conference on Advances in Computer Applications (ICACA)*, 2016, pp. 235–238.
- [92] A. Mahgoub, S. Ganesh, F. Meyer, A. Grama, and S. Chaterji, “Suitability of nosql systems — cassandra and scylladb — for IoT workloads,” in *2017 9th International Conference on Communication Systems and Networks (COM-SNETS)*, Jan 2017, pp. 476–479.
- [93] P. Patel, M. Intizar Ali, and A. Sheth, “On using the intelligent edge for IoT analytics,” *IEEE Intelligent Systems*, vol. 32, no. 5, pp. 64–69, Sep. 2017.
- [94] M. I. Naas, P. R. Parvedy, J. Boukhobza, and L. Lemarchand, “iFogStor: An IoT data placement strategy for fog infrastructure,” in *IEEE 1st International Conference on Fog and Edge Computing (ICFEC)*, May 2017, pp. 97–104.
- [95] A. Aral and T. Ovatman, “A decentralized replica placement algorithm for edge computing,” *IEEE Transactions on Network and Service Management*, vol. 15, no. 2, pp. 516–529, June 2018.

- [96] Y. Shao, C. Li, and H. Tang, "A data replica placement strategy for IoT workflows in collaborative edge and cloud environments," *Computer Networks*, vol. 148, 11 2018.
- [97] W. Wang, G. Wu, Z. Guo, L. Qian, L. Ding, and F. Yang, "Data scheduling and resource optimization for fog computing architecture in industrial IoT," *15th International Conference, ICDCIT*, pp. 141–149, 01 2019.
- [98] K. Velasquez, D. P. Abreu, M. Curado, and E. Monteiro, "Service placement for latency reduction in the internet of things," *Annals of Telecommunications*, vol. 72, pp. 105–115, Feb. 2017.
- [99] L. Sun, Y. Li, and R. A. Memon, "An open IoT framework based on microservices architecture," *China Communications*, vol. 14, no. 2, pp. 154–162, February 2017.
- [100] N. Mohamed, J. Al-Jaroodi, I. Jawhar, S. Lazarova-Molnar, and S. Mahmoud, "Smartcityware: A service-oriented middleware for cloud and fog enabled smart city services," *IEEE Access*, vol. 5, pp. 17 576–17 588, 2017.
- [101] U. Ozeer, X. Etchevers, L. Letondeur, F.-G. Ottogalli, G. Salaün, and J.-M. Vincent, "Resilience of stateful IoT applications in a dynamic fog environment," in *Proceedings of the 15th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. ACM, nov 2018.
- [102] A. Giordano, G. Spezzano, and A. Vinci, "Smart agents and fog computing for smart city applications," in *International Conference on Smart Cities*, 06 2016, pp. 137–146.
- [103] B. Tang, Z. Chen, G. Heffernan, T. Wei, H. He, and Q. Yang, "A hierarchical distributed fog computing architecture for big data analysis in smart cities," in *ASE International Conference on Big Data*, 10 2015.
- [104] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," *CoRR*, vol. abs/1610.05492, 2016. [Online]. Available: <http://arxiv.org/abs/1610.05492>
- [105] B. McMahan and D. Ramage, "Federated learning: Collaborative machine learning without centralized training data," Google, Tech. Rep., 2017.
- [106] U. Roy, "Leak detection in pipe networks using hybrid ann method," *Water Conservation Science and Engineering*, vol. 2, no. 4, p. 145–152, 2017.

- [107] K. S. Dar, A. Taherkordi, and F. Eliassen, “Enhancing dependability of cloud-based IoT services through virtualization,” in *2016 IEEE First International Conference on Internet-of-Things Design and Implementation (IoTDI)*, April 2016, pp. 106–116.
- [108] H. Yu and M. Guo, “An efficient oil and gas pipeline monitoring systems based on wireless sensor networks,” in *International Conference on Information Security and Intelligent Control*, Aug 2012, pp. 178–181.
- [109] A. Ioana, C. Burlacu, and A. Korodi, “Approaching OPC UA publish–subscribe in the context of UDP-based multi-channel communication and image transmission,” *Sensors*, vol. 21, no. 4, p. 1296, feb 2021.
- [110] A. Aslam and E. Curry, “Investigating response time and accuracy in online classifier learning for multimedia publish-subscribe systems,” *Multimedia Tools and Applications*, jan 2021.
- [111] H. Jafarpour, B. Hore, S. Mehrotra, and N. Venkatasubramanian, “CCD: A distributed publish/subscribe framework for rich content formats,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 5, pp. 844–852, may 2012.
- [112] D. Garg, N. C. Narendra, and S. Tesfatsion, “Heuristic and reinforcement learning algorithms for dynamic service placement on mobile edge cloud,” *CoRR*, 2021.
- [113] T. He, H. Khamfroush, S. Wang, T. La Porta, and S. Stein, “It’s hard to share: Joint service placement and request scheduling in edge clouds with sharable and non-sharable resources,” in *2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*, 2018, pp. 365–375.
- [114] Q. Cai, W. Hang, A. Mirhoseini, G. Tucker, J. Wang, and W. Wei, “Reinforcement learning driven heuristic optimization,” *CoRR*, 2019.
- [115] M. S. Islam, R. Nix, and M. Kantarcioglu, “A game theoretic approach for adversarial pipeline monitoring using wireless sensor networks,” in *2012 IEEE 13th International Conference on Information Reuse Integration (IRI)*, 2012, pp. 37–44.
- [116] A. Rezazadeh, L. Talarico, G. Reniers, V. Cozzani, and L. Zhang, “Applying game theory for securing oil and gas pipelines against terrorism,” *Reliability Engineering & System Safety*, vol. 191, p. 106140, nov 2019.

- [117] A. Rezazadeh, L. Zhang, G. Reniers, N. Khakzad, and V. Cozzani, “Optimal patrol scheduling of hazardous pipelines using game theory,” *Process Safety and Environmental Protection*, vol. 109, pp. 242–256, jul 2017.
- [118] S. Ahmed, F. Le Mouël, and N. Stouls, “Resilient IoT-based Monitoring System for Crude Oil Pipelines,” in *Proceedings of the 7th International Conference on Internet of Things: Systems, Management and Security (IOTSMS)*. IEEE, Dec. 2020.
- [119] “Iot device standards,” in *2014 IEEE Hot Chips 26 Symposium (HCS)*, 2014, pp. 1–37.
- [120] K. Mekki, E. Bajic, F. Chaxel, and F. Meyer, “A comparative study of LPWAN technologies for large-scale IoT deployment,” *ICT Express*, vol. 5, no. 1, pp. 1–7, mar 2019.
- [121] F. Ciccozzi, I. Crnkovic, D. Di Ruscio, I. Malavolta, P. Pelliccione, and R. Spalazzese, “Model-driven engineering for mission-critical IoT systems,” *IEEE Software*, vol. 34, no. 1, pp. 46–53, Jan 2017.
- [122] N. Matni, Y. P. Leong, Y. S. Wang, S. You, M. B. Horowitz, and J. C. Doyle, “Resilience in large scale distributed systems,” *Procedia Computer Science*, vol. 28, pp. 285–293, 2014.
- [123] M. Correia, N. F. Neves, and P. Verissimo, “How to tolerate half less one byzantine nodes in practical distributed systems,” in *Proceedings of the 23rd IEEE International Symposium on Reliable Distributed Systems, 2004*. IEEE, 2004, pp. 174–183.
- [124] F. Cappello, A. Geist, B. Gropp, L. Kale, B. Kramer, and M. Snir, “Toward exascale resilience,” *The International Journal of High Performance Computing Applications*, vol. 23, no. 4, pp. 374–388, 2009.
- [125] D. Liu, R. Deters, and W.-J. Zhang, “Architectural design for resilience,” *Enterprise Information Systems*, vol. 4, no. 2, pp. 137–152, 2010.
- [126] K. M. Alam, J. Kamruzzaman, G. Karmakar, and M. Murshed, “Dynamic adjustment of sensing range for event coverage in wireless sensor networks,” *Journal of Network and Computer Applications*, vol. 46, pp. 139–153, nov 2014.
- [127] G. Solmaz and D. Turgut, “Optimizing event coverage in theme parks,” *Wireless networks*, vol. 20, no. 6, pp. 1445–1459, 2014.

- [128] M. Yousef, K. F. Hussain, and U. S. Mohammed, “Accurate, data-efficient, unconstrained text recognition with convolutional neural networks,” *Pattern Recognition*, vol. 108, p. 107482, dec 2020.
- [129] Y. Wu and Y. Lin, “Inversionnet: An efficient and accurate data-driven full waveform inversion,” *IEEE Transactions on Computational Imaging*, vol. 6, pp. 419–433, 2020.
- [130] S. Ahmed, F. Le Mouël, N. Stouls, and G. Lipeme Kouyi, “Hydillech: a WSN-based Distributed Leak Detection and Localisation in Crude Oil Pipelines,” in *Proceedings of the 45th International Conference on Advanced Information Networking and Applications (AINA-2020)*. IEEE, oct 2021.
- [131] T. Braun, A. Kassler, M. Kihl, V. Rakocevic, V. Siris, and G. Heijenk, *Traffic and QoS Management in Wireless Multimedia Networks*. Springer Verlag, USA, 2009, vol. 31.
- [132] G. Parissidis, M. Karaliopoulos, M. May, T. Spyropoulos, and B. Plattner, “Interference in wireless multihop networks: A model and its experimental evaluation,” in *International Symposium on a World of Wireless, Mobile and Multimedia Networks*, 2008, pp. 1–12.
- [133] G. Klimiashvili, C. Tapparello, and W. Heinzelman, “Lora vs. wifi ad hoc: A performance analysis and comparison,” in *2020 International Conference on Computing, Networking and Communications (ICNC)*, 2020, pp. 654–660.
- [134] M. Mohammadpoor and F. Torabi, “Big data analytics in oil and gas industry: An emerging trend,” *Petroleum*, 12 2018.
- [135] R. Aliguliyev and Y. Imamverdiyev, “Conceptual big data architecture for the oil and gas industry,” *Problems of Information Technology*, vol. 08, pp. 3–13, 01 2017.
- [136] S. Ahmed, F. Le Mouël, N. Stouls, and J. S. Dibanoye, “R-MDP: A Game Theory Approach for Fault-Tolerant Data and Service Management in Crude Oil Pipelines Monitoring Systems,” in *Proceedings of the 19th International Conference on Mobile Information Networking and Applications (MobiQuitous-2022)*. Springer, Nov. 2022.
- [137] S. Porkodi and D. Kesavaraja, “Integration of blockchain and internet of things,” in *Handbook of Research on Blockchain Technology*. Elsevier, 2020, pp. 61–94.

- [138] P. Zhang, D. C. Schmidt, J. White, and A. Dubey, "Consensus mechanisms and information security technologies," in *Advances in Computers*. Elsevier, 2019, pp. 181–209.
- [139] X. Wang, X. Zha, W. Ni, R. P. Liu, Y. J. Guo, X. Niu, and K. Zheng, "Survey on blockchain for internet of things," *Computer Communications*, vol. 136, pp. 10–29, feb 2019.
- [140] Y. Shoham and K. Leyton-Brown, *MULTIAGENT SYSTEMS Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, 2009,2010.
- [141] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, second edition ed. The MIT Press, 2020.
- [142] P. Gawlowicz and A. Zubow, "ns-3 meets openai gym: The playground for machine learning in networking research," in *ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM)*, Nov. 2019.
- [143] D. Garg, N. C. Narendra, and S. Tesfatsion, "Heuristic and reinforcement learning algorithms for dynamic service placement on mobile edge cloud," 2021.
- [144] D. Magrin, M. Centenaro, and L. Vangelista, "Performance evaluation of LoRa networks in a smart city scenario," in *2017 IEEE International Conference on Communications (ICC)*. IEEE, may 2017.
- [145] M. Capuzzo, D. Magrin, and A. Zanella, "Confirmed traffic in LoRaWAN: Pitfalls and countermeasures," in *2018 17th Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net)*. IEEE, jun 2018.
- [146] K. Shan and J. Shuai, "Statistical analyses of incidents on oil and gas pipelines based on comparing different pipeline incident databases," in *Volume 3B: Design and Analysis*. American Society of Mechanical Engineers, jul 2017.
- [147] Y. Gao, "Two players game based on schrödinger equation solution," *Theoretical Economics Letters*, vol. 12, no. 02, pp. 564–574, 2022.
- [148] J. Bonvoisin, A. Lelah, F. Mathieux, and D. Brissaud, "An environmental assessment method for wireless sensor networks," *Journal of Cleaner Production*, vol. 33, pp. 145–154, sep 2012.

- [149] —, “An integrated method for environmental assessment and ecodesign of ICT-based optimization services,” *Journal of Cleaner Production*, vol. 68, pp. 144–154, apr 2014.
- [150] J. Pohl, L. Hilty, and M. Finkbeiner, “How lca contributes to the environmental assessment of higher order effects of ict application: A review of different approaches,” *Journal of Cleaner Production*, 2019.
- [151] M. A. Achachlouet, “Exploring the Effects of ICT on Environmental Sustainability: From Life Cycle Assessment to Complex Systems Modelling,” Ph.D. dissertation, Royal Institute of Technology, 2015.
- [152] “Oil and petroleum products explained,” *U.S. Energy Information Administration*, 2022.
- [153] T. He, Z. R. Chong, J. Zheng, Y. Ju, and P. Linga, “LNG cold energy utilization: Prospects and challenges,” *Energy*, vol. 170, pp. 557–568, mar 2019.
- [154] “Liquefied natural gas,” *Fossil Energy and Carbon Management*, 2022.
- [155] G. Molnar, “Economics of gas transportation by pipeline and lng,” in *The Palgrave Handbook of International Energy Economics*. Springer, 2022, pp. 23–57.
- [156] C. E. Gómez-Camacho, B. Ruggeri, and R. Hirsch, “An environmental perspective on natural gas transport options: Pipelines vs liquefied natural gas (lng),” *SUSTAINABLE ENERGY AND ENVIRONMENTAL PROTECTION*, 2022.
- [157] A. C. J. Malar, C. A. Glover, S. R. Prasath, A. P. Parameshwari, R. R. Rekanivetha, R. Sarveshwaran, and R. T. P. Kumar, “Smart and innovative water conservation and distribution system for smart cities,” in *IoT and IoE Driven Smart Cities*. Springer International Publishing, dec 2021, pp. 253–273.
- [158] K. B. Adediji, A. A. Ponnle, A. M. Abu-Mahfouz, and A. M. Kurien, “Towards digitalization of water supply systems for sustainable smart city development—water 4.0,” *Applied Sciences*, vol. 12, no. 18, p. 9174, 2022.
- [159] A. Hassanin, M. Mir, and F. Mukhtar, “A wireless sensor network for water pipeline leak detection,” in *Water, Energy and Environment 2011*, p. 143, 2011.

- [160] M. Y. Aalsalem, W. Z. Khan, W. Gharibi, and N. Armi, “An intelligent oil and gas well monitoring system based on internet of things,” in *2017 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET)*, 2017, pp. 124–127.
- [161] T. R. Wanasinghe, R. G. Gosine, L. A. James, G. K. Mann, O. De Silva, and P. J. Warrian, “The internet of things in the oil and gas industry: a systematic review,” *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8654–8673, 2020.
- [162] F. Sattari, L. Lefsrud, D. Kurian, and R. Macciotta, “A theoretical framework for data-driven artificial intelligence decision making for enhancing the asset integrity management system in the oil and gas sector,” *Journal of Loss Prevention in the Process Industries*, vol. 74, p. 104648, jan 2022.