N° d'ordre 391/21 Année 2021

UNIVERSITÉ SULTAN MOULAY SLIMANE
Faculté des Sciences et Techniques Béni-Mellal

جامعة السلطان مولاي سليمان
Université Sultan Moulay Slimane

FST - BM

Centre d'Etudes Doctorales : « Sciences et Techniques»
Formation Doctorale : « Mathématiques et Physique Appliquées»

THESE

Présentée Par

IFZARNE SAMIR

Pour l'obtention du grade de

DOCTEUR

Discipline : Informatique

Spécialité : Sécurité et Machine Learning

GESTION DE LA SECURITE DANS LES RESEAUX DE CAPTEURS SANS FIL VERS L'APPROCHE DE PROTECTION INTELLIGENTE

Soutenue le 28 Décembre 2021 devant le Jury

Président :
Pr. Abdelmoutalib METRANE (Ecole Nationale des Sciences Appliquées de Khouribga)
Rapporteurs :
Pr. Noureddine ABOUTABIT (Ecole Nationale des Sciences Appliquées de Khouribga)
Pr. Said EL KAFHALI (Faculté des Sciences et Techniques de Settat)
Pr. Abdelmoutalib METRANE (Ecole Nationale des Sciences Appliquées de Khouribga)
Examinateurs :
Pr. Abdelghani GHAZDALI (Ecole Nationale des Sciences Appliquées de Khouribga)
Co-Encadrant :
Pr. Nadia IDRISSI FATMI (Ecole Nationale des Sciences Appliquées de Khouribga)
Directeur de thèse :
Pr. Imad HAFIDI (Ecole Nationale des Sciences Appliquées de Khouribga)
Invité :
Dr. Younes BENCHIGRA (Maroc Telecom)

Structure de recherche accréditée d'accueil :
USMS/ENSA/MAI : Laboratoire Ingénierie des Procédés, Informatique et Mathématique (LIPIM) de l'ENSA de Khouribga

Acknowledgement

I owe my deepest gratitude to my advisors Prof. Imad HAFIDI and Prof. Nadia IDRISSI FATMI for their trust, continuous support, relentless and insightful guidance. Thank you, your motivation, knowledge and long lasting experience were irreplaceable assets for me.

Many thanks to my brother Pr. Aziz IFZARNE who has been putting me into the research path. Even though I may not say it all the time, I appreciate so much all what he has done since day 0 until completion.

I am deeply indebted to my Friend Dr. Younes BENCHIGRA for his helpful advice and useful guidance related to my research. Without your help, I wouldn't have finished this thesis.

I would like to thank all members of Process engineering, Computer science and Mathematics Laboratory (LIPIM: Laboratoire Ingénierie des Procédés, Informatique et Mathématique) for their kindness and advises.

A special thanks to my family, words cannot express how grateful I am to my wife, my daughter and my son for all of the sacrifices they made during the time I spent to accomplish this work.

Dedication

I would like to dedicate my work to

My father and my mother

My step-parents

My wife

My daughter Lamisse and my son Nawrasse

My brothers and my cousins

My friends and colleagues

ABSTRACT

Wireless sensor network (WSN) plays a vital role in environment monitoring and various other applications like precision agriculture or buildings structure monitoring. Generally, network performance relies on the capacity of sensor nodes to process and deliver data packets to the base station at lower cost and with the required security level. One of the imperative features of WSN is providing secure data protection since sensor nodes sends their readings within the network until reaching the base station. However, security requires high computation and additional data communication by sensor nodes, which get data delivery process costing more energy, and network lifetime reduced and related applications do not work properly. In order to guarantee secure data packet delivery in respect of expected data protection requirements, numerous security schemes has been proposed attempting to protect the data and detect attacks on sensors and network. Unfortunately, proposed security schemes does not protects against all network attacks and each one just protect data confidentiality against specific attacks or under specific network conditions, such as low dense network size. Therefore, confidentiality and attack detection in WSN should be addressed carefully while designing security schemes.

In this thesis, we developed two novel cluster based secure data aggregation (CSDA) schemes respectively (CSHEAD) and (PC2SR) to protect data confidentiality and detect attacks during aggregation by cluster head. The design and implementation of proposed CSHEAD and PC2SR is based on a combination of semi-homomorphic encryption and compressive sensing techniques. Encryption is ensuring data protection and homomorphic property enables applying arithmetic operations on cipher text without losing capability to decrypt. While this is combined with compressive sensing, data volume reduces significantly and thus energy consumption. Also the compression is distributed to every sensor node at low computation cost. The second contribution of this thesis is the proposition of an attack detection scheme named Intrusion Detection - Gain ratio Online Progressive Aggressive (ID-GOPA) algorithm. ID-GOPA is using machine learning techniques to detect several types of attacks in wireless sensor network. Performance evaluation of proposed CSHEAD, PC2SR and ID-GOPA has been conducted based on simulation in NS2. Performance analysis based on simulation outputs has proved that CSHEAD and PC2SR outperforms CSDA in terms of energy consumption, attack detection, Packet delivery Ratio, Throughput and routing overhead. Second set of simulations between ID-GOPA and other methods of WSN intrusion detection models like SVM, Naïve Bayes, Random Forest and Decision Tree shows that ID-GOPA has better detection accuracy, Recall, Precision and better F1-score.

***Keywords**: Wireless Sensor Network (WSN), Compressive Sensing, Homomorphic Encryption, Secure Data Aggregation, Machine Learning, Intrusion Detection, Online Learning.*

RÉSUMÉ

Le réseau de capteurs sans fil (WSN) joue un rôle essentiel dans la surveillance de l'environnement et diverses autres applications telles que l'agriculture de précision ou la surveillance de la structure des bâtiments. Généralement, les performances du réseau reposent sur la capacité des nœuds capteurs à traiter et livrer des paquets de données à la station de base à moindre coût et avec le niveau de sécurité requis. L'une des caractéristiques impératives de WSN est de fournir une protection et sécurité des données tandis que les nœuds de capteurs envoient leurs lectures au sein du réseau jusqu'à atteindre la station de base. Cependant, la sécurité nécessite un calcul élevé et une communication de données supplémentaire par les nœuds de capteurs, ce qui entraîne un processus de livraison de données coûtant plus d'énergie, et la durée de vie du réseau est réduite et les applications associées ne fonctionnent pas correctement. Afin de garantir une livraison sécurisée des paquets de données en respectant les exigences de protection des données attendues, de nombreux schémas de sécurité ont été proposés pour tenter de protéger les données et détecter les attaques sur les capteurs et le réseau. Malheureusement, les schémas de sécurité proposés ne protègent pas contre toutes les attaques de réseau et chacun protège simplement la confidentialité des données contre des attaques spécifiques ou dans des conditions de réseau spécifiques, telles qu'une taille de réseau faiblement dense. Par conséquent, la confidentialité et la détection des attaques dans WSN doivent être traitées avec soin lors de la conception des schémas de sécurité.

Dans cette thèse, nous avons développé deux nouveaux schémas d'agrégation sécurisée de données (CSDA) basés sur des clusters respectivement (CSHEAD) et (PC2SR) pour protéger la confidentialité des données et détecter les attaques lors de l'agrégation par cluster Head. La conception et la mise en œuvre des CSHEAD et PC2SR proposés sont basées sur une combinaison de techniques de cryptage semi-homomorphe et de détection compressive. Le chiffrement garantit la protection des données et la propriété homomorphe permet d'appliquer des opérations arithmétiques sur le texte chiffré sans perdre la capacité de déchiffrement. Bien que cela soit combiné à la détection compressive, le volume de données est réduit considérablement et ainsi pareil pour la consommation d'énergie. De plus, la compression est distribuée à chaque nœud de capteur et a un faible coût de calcul. La deuxième contribution de cette thèse est la proposition d'un schéma de détection d'attaque nommé Intrusion Detection - Gain ratio Online Progressive Aggressive (ID-GOPA). ID-GOPA utilise des techniques d'apprentissage automatique pour détecter plusieurs types d'attaques dans un réseau de capteurs sans fil. L'évaluation des performances des propositions CSHEAD, PC2SR et ID-GOPA a été réalisée sur la base d'une simulation dans NS2. L'analyse des performances basée sur les sorties de simulation a prouvé que CSHEAD et PC2SR surpassent CSDA en termes de consommation d'énergie, de détection d'attaque, de taux de livraison de paquets, de débit et de surcharge de routage. La deuxième série de simulations entre ID-GOPA et d'autres méthodes de détection d'intrusion WSN comme SVM, Naïve Bayes, Random Forest et Decision Tree montre que ID-GOPA a une meilleure précision de détection, Rappel, Précision et un meilleur score F1.

***Mots clés** : Réseau de capteurs sans fil (WSN), détection compressive, chiffrement homomorphe, agrégation sécurisée de données, apprentissage automatique, détection d'intrusion, apprentissage en ligne.*

Table of Contents

List of Figures

List of Tables

List of publications

Journal articles

1. Ifzarne, S., Hafidi, I. & Idrissi, N. Compressive sensing and paillier cryptosystem based secure data collection in WSN. J Ambient Intell Human Comput (September 2021), doi: https://doi.org/10.1007/s12652-021-03449-6.
2. Samir Ifzarne, Imad Hafidi, and Nadia Idrissi, "A Novel Secure Data Aggregation Scheme Based on Semi-Homomorphic Encryption in WSNs," Journal of Communications vol. 16, no. 8, pp. 323-330, August 2021. doi: https://doi.org/10.12720/jcm.16.8.323-330.
3. S. Ifzarne, H. Tabbaa, H. Imad, and N. Lamghari, “Anomaly Detection using Machine Learning Techniques in Wireless Sensor Networks,” Journal of Physics: Conference Series, vol. 1743, p. 012021, Jan. 2021, doi: https://doi.org/10.1088/1742-6596/1743/1/012021.

Conference papers

1. S. Ifzarne, H. Imad, and N. Idrissi, “Compressive Sensing Based on Homomorphic Encryption and Attack Classification using Machine Learning Algorithm in WSN Security,” NISS2020: Proceedings of the 3rd International Conference on Networking, Information Systems & Security, March 2020, Article No.: 40, Pages 1–6, doi: https://doi.org/10.1145/3386723.3387859.
2. S. Ifzarne, H. Imad, and N. Idrissi, “Homomorphic Encryption for Compressed Sensing in Wireless Sensor Networks,” SCA '18: Proceedings of the 3rd International Conference on Smart City Applications, October 2018 Article No.: 80, Pages 1–6, doi: https://doi.org/10.1145/3286606.3286857.

Book chapters

1. S. Ifzarne, H. Imad, and N. Idrissi, “Data Aggregation Privacy in WSN Combined with Compressive Sensing”, Special Issue on Data and Security Engineering”, SCA: Innovations in Smart Cities Applications Edition 2 2019, pp. 691–701. doi: https://doi.org/10.1007/978-3-030-11196-0_57

2. Ifzarne S., Hafidi I., Idrissi N. (2021) Secure Data Collection for Wireless Sensor Network. In: Ben Ahmed M., Mellouli S., Braganca L., Anouar Abdelhakim B., Bernadetta K.A. (eds) Emerging Trends in ICT for Sustainable Development. Advances in Science, Technology & Innovation (IEREK Interdisciplinary Series for Sustainable Development). Springer, Cham. https://doi.org/10.1007/978-3-030-53440-0_26

Glossary of terms

1G	First mobile networks generation
2G	Second mobile networks generation
3G	Third mobile networks generation
3GPP	Generation Partnership Project
4G	Fourth mobile networks generation
5G	Fifth mobile networks generation
ABC	Artificial Bee Colony
ABR	Associativity Based Routing
ACO	Ant Colony Optimization
ADC	Analog to Digital Converter
AI	Artificial Intelligence
ALOHA	Areal Location of Hazardous Atmospheres
ANN	Artificial Neural Network
AODV	Ad hoc On demand Distance Victor routing
AODVM	Multipath Ad hoc On demand Distance Victor routing
API	Application Programming Interface
B.C	Before Christ
BS	Base Station
CCIPCA	Candid Covariance free Incremental Principal Component Analysis
CDMA	Code Division Multiple Access
CO	Carbon Monoxide
CP	Constraint Programming
CSA	Cuckoo Search Algorithm
CSDA	Cluster-based Secure Data Aggregation
CSHEAD	Cluster-based Semi-Homomorphic Encryption Aggregated Data
CSMA	Carrier Sense Multiple Access
CSMA-CA	Carrier Sense Multiple Access with Collision Avoidance
DARPA	Defense Advanced Research Project Agency
DOD	Department Of Defense
DREAM	Distance Routing Effect Algorithm For Mobility

DSDV	Destination-Sequenced Distance Victor routing
DSL	Digital Subscriber Line
DSR	Dynamic Source Routing
DT	Decision Tree
DVR	Distance Victor Routing
ECG	ElectroCardioGram
EM	Electromagnetic
EP	Evolutionary Programming
ES	Evolution Strategy
ETT	Expected Transmission Time
ETX	Expected Transmission Count
FANET	Flaying Ad hoc Network
FHSS	Frequency Hop Spread Spectrum
FIRMS	Fire Information for Resource Management System
FSR	Fisheye Sate Routing
GloMo	Globe Mobile information system
GM	Gauss Markov
GNM	Global Normal Model
GP	Genetic Programming
IARP	Intra-zone Routing
ID-GOPA	Intrusion Detection - Gain ratio Online Progressive Aggressive
IDS	Intrusion Detection System
IEEE	Institute of Electrical and Electronics Engineers
IERP	Extra-zone Routing
IoT	Internet of Things
IP	Internet Protocol
ISIS	Intermediate System to Intermediate System
ISP	Internet Service Provider
LAN	Local Area Network
LIP	Linear Integer Programming
LSP	Link State Packet
LSR	Link State Routing
LSSVM	Least Squares-Support Vector Machine
LTE	Long Term Evolution
LEACH	Low-energy adaptive clustering hierarchy
MAC	Media Access Control

MAN	Metropolitan Area Network
MCN	Multi-hop Cellular Network
MCR	Multi-Channel Routing metric
MIP	Mixed Integer Programming
MMBCR	Min-Max Battery Cost Routing
MN	Mobile Node
MTPR	Minimal Total Power Routing
NB	Naïve Bayes
NLP	Non Linear Programming
NO2	Nitrogen Dioxide
NS2	Network simulator 2
NTDR	Near Term Digital Radio
OCPCC	One-Class Principal Component Classifier
OLSR	Optimized Link State Routing
OMRP	On demand Multicast Routing Protocol
OSPF	Open Short Path First
OTCL	Object oriented Tool Command Language
P2P	Peer To Peer
PA	Passive-Aggressive Algorithm
PAN	Personal Area Network
PC2SR	Paillier Cryptosystem and Compressive Sensing based Routing
PCF	Path Change Factor
PCMCIA	Personal Computer Memory Card International Association
PDA	Personal Digital Assistants
PDR	Packet Delivery Ratio
PRNET	Packet Radio Network
QoS	Quality of Service
QoSR	Quality of Service Routing
RABR	Route lifetime Assessment Based Routing
RBF	Radial Basis Function
RERR	Route Error
RF	Random Forest
RIP	Routing Information Protocol
RKHS	Reproducing Kernel Hilbert Space
RREP	Route Response
RREQ	Route Request

RRI	Route repair influence
RLS	Recursive Least Squares
RW	Random Walk
RWP	Random Way Point
SAMCP	Simulated Annealing Multi-constrained Path
SA	Simulated Annealing
SC	Switching Cost
SDN	Software Defined Network
SI	Swarm Intelligence
SLR	Service Level Requirement
SONET	Synchronous Optical Networking
SOPRANO	Self-Organizing Packet Radio Ad hoc Network with Overlay
SSAR	Signal Stability based Adaptive Routing
SURAN	Survivable Adaptive Radio Network
SVM	Support Vector Machine
TDMA	Time Division multiple Access
TS	Tabu Search
TTL	Time To Live
UANET	Underwater Ad hoc Network
UAV	Unmanned Air Vehicles
VANET	Vehicular Ad hoc Network
VOIP	Voice Over Internet Protocol
VRP	Vehicle Routing Problem
WAN	Wide Area Network
WCETT	Weighted Cumulative Expected Transmission Time
WDM	Wavelength Division Multiplexing
Wi-Fi	Wireless Fidelity
WiMax	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Network
WMN	Wireless Mesh Network
WPAN	Wireless Personal Area Network
WSN	Wireless Sensor Network
ZHLS	Zone-Based Hierarchical Link State routing

General Introduction

1. Scope of work

Since last few years, the wireless industry has been growing at an exponential rate, shaping human's ecosystem and making its life easier. This is noticeable in the way people interact, shop and do business. Earlier, communications were essentially based on landlines; fax machines and physical mail. Nowadays, due to mobile computing advances, the wide availability of wireless networking facilities and the emergence of handheld computers, personal digital assistants (PDAs) and smart phones; we can easily achieve things that were unbelievable in the near past. For example, monitor and control connected devices in the home network; keep track of pressure inside pipelines for Oil and Gas in real time; consult agriculture field's and crops information's like temperature and humidity allowing farmers to make right decisions to optimize the usage of water and fertilizer and also maximize the yield of the crops;. We can also prevent adverse consequences and pro-actively manage natural disasters like floods where changes in rivers water levels can be monitored in real time; Health care monitoring is another example where devices embedded in the body environment can track the physical state of a person for continuous health diagnosis making health care more efficient and less expensive as continuous monitoring can be done remotely while the patient doesn't require to stay in a healthcare center.

All this ease is the result of low cost tiny wireless sensor devices and the ability to develop an on-demand, self-organizing wireless network without relying on any available fixed infrastructure (called wireless sensor network) which can also interface with the classic wireless network based on fixed infrastructure (called *infrastructure-based wireless network*). Wireless sensor networks (WSNs) are cost effective and easy to deploy. This type of network represents an alternative of well-known wired sensors in the areas where the infrastructure does not exist or needs too much time and high investment to build it up. A typical application of WSNs can be described as a group of firefighting team in a forest under fire. Sensors deployed prior to the fire or quickly from the air after the fire started can wirelessly connect to each other with the help of efficient routing protocols that help them to self-organize and maintain quality of communication and start sending temperature figures and their positions to fire brigade to know how the fire is spreading and how to manage it efficiently. WSN performance depends essentially on the application requirements which are related to network capability of delivering sensor readings from sensor nodes to the base station while taking care of data privacy protection especially for critical applications like battlefield monitoring. However, network security efficiency may be drastically affected by the computational cost needed by encryption cryptosystems and the limited power based on batteries for the nodes lifetime. Therefore, the main aim of our work is to explore current research state in data protection and attack detection then propose novel efficient approach that could deliver better performance while managing security requirements.

2. Problem statement and research gap

Data protection in WSN plays an important role in security schemes, where data confidentiality and integrity should be managed while each node acts as router in addition of being source or destination node. The wireless nature of the communication means all communication packets are accessible for eavesdroppers which can even read data if not encrypted. Also, nodes are generally deployed in harsh environment making them physically exposed to capture by adversaries. The main issue that affects network security in WSNs is the limited resources of energy, memory and computation of sensor nodes. Each sensor node within the network should encrypt data and participate to attack detection as well as other network operations. Thus, the secure delivery of data packets to the base station is heavily affected and pose vital challenging problems. Numerous data protection schemes have been proposed to search for an appropriate encryption and key distribution model that protect data confidentiality without increasing calculation complexity, volume of data and wireless communication. The main aim is to establish or build an encrypted communication channel between network nodes. However, up to now there has not been an optimal cryptosystem that produces the encryption and decryption without affecting network performance since most of available encryption model were initially designed for systems with no resource constraints or modified for wireless sensor network. Also, attack detection algorithms have been developed to spot specific attacks. As long as the number of attacks is increasing and attackers in general don’t have resource constraint issue, building an attack detection system able to perceive all attacks was not possible. In addition to energy consumption by wireless communication, some other challenges must be taken in account during the design of attack detection schemes for WSNs. For instance, the broadcast nature of the wireless medium introduces the hidden-terminal and exposed-terminal problems; the variable and unpredictable capacity of wireless links; restricted power nodes; computing capacity and scarce bandwidth.

Numerous data protection and attack detection strategies have been adopted. The First strategy named secure data aggregation consist of monitoring the operations of Cluster nodes as they are the most critical nodes and aggregation is a high critical step in data forwarding. Thus, each node send parts of its encrypted readings to cluster member's and receives the same from the other nodes, which permits all nodes within the same cluster to aggregate data from same round and control whether the cluster head is sending the same data to next hop or to the base station. However, keeping wireless communication active for data reception by all nodes consumes a lot of energy. This generate a high communication overhead that will decrease network performance. The second strategy is attack detection that are designed to discover the known attacks based on offline machine learning algorithms. Consequently, the learning period is scheduled in the first step before the network start its operations and needs a periodical update to get the new attacks signatures. This attach detection is quit adapted to WSN environment as it's not developed for a specific attack but for a set of various known attacks. The third approach is hybrid data protection and attack detection. It combine the advantages of both encryption to protect data and attack detection to achieve less overhead as in separated approach for confidentiality and attack detection. The hybrid mode is valid only for specific attack detection algorithms and not adapted to machine learning algorithms. However, performance comparison and evaluation of encryption and attack detection under different contexts established the fact that there is no optimal encryption scheme that provides better network performance within all contexts.

3. Thesis contribution

Throughout extensive literature review, we conducted a conceptual study based on working process of various pre-established secure data aggregation protocols and intrusion detection systems. We investigated the combination of compressive sensing and semi-homomorphic encryption for data protection in hierarchical routing protocols. The outcomes of this study has led to suppose that asymmetric encryption can be managed when reducing the data volumes in a distributed compression. To confirm that hypothesis, we simulated three secure data aggregation schemes, CSDA, CSHEAD and PC2SR using NS2. The first contribution is the proposition of two new schemes; Cluster-based Semi-Homomorphic Encryption Aggregated Data (CSHEAD) and Paillier Cryptosystem and Compressive Sensing based Routing (PC2SR) to improve network performance while managing data confidentiality and attack detection in the aggregation phase. The second contribution is the development of novel model for intrusion detection system called ID-GOPA, which is based on a combination of offline learning and online machine learning for anomalies detection algorithms. The proposed ID-GOPA stands for Intrusion Detection - Gain ratio Online Progressive Aggressive model which uses offline learning for feature selection and building initial trained model, then the online learning takes over to continuously update the model based on online sensor and network data. First contribution performance has been evaluated and comparison based on six metrics such delay, throughput, routing overhead, PDR, attack detection accuracy and energy consumption has confirmed that both CSHEAD and PC2SR outperforms CSDA. Evaluation of the proposed intrusion detection system was done against SVM, NB, RF, and DT using the same dataset (WSN-DS). The performance comparison was done using four metrics namely Recall, precision, F1-score and accuracy index. Our incremental machine learning approach ID-GOPA deliver better performance results overall.

4. Thesis organization

The present thesis starts with the introduction where we present the scope of work, problem statement, research gap and thesis contribution followed by four main chapters and ended by a conclusion & perspectives.

The first chapter presents an overview of Wireless sensor networks. First we pointed out different types of existing networks as well as infrastructure and infrastructure less based networks and their evolution. A focus on wireless sensor network (WSN) evolution, challenges and application is done in the rest of this chapter.

The second chapter is dedicated to review the state of the art in WSN security, establish taxonomy of data aggregation and review routing protocols in WSNs with a focus on LEACH protocol and study the compressive sensing. Security requirements in the context of WSN as well as attacks classification have been presented. Link between different techniques is established based on the fact that security is a must for WSN adoption and scare energy, computational resources as a constraint that can’t be modified but should be managed.

The third chapter is devoted to present the two proposed secure data aggregation schemes based on homomorphic encryption, compressive sensing and LEACH protocol. CSHEAD Protocol is an improvement of the existing CSDA where all nodes in a cluster contribute to monitor cluster head operations. The chapter present as well design for second proposal namely PC2SR protocol followed by an evaluation to validate stated hypothesis via simulating CSDA, CSHEAD and PC2SR using multiple value of network density by changing number of nodes within same area. Performance evaluation were carried out based on four traditional metrics such as delay, throughput, PDR and routing overhead and two additional metrics energy consumption and attack detection accuracy that reflect the effectiveness of secure data aggregation in WSN context. I addition we motivate the choice of NS2 simulator. In the fourth chapter, we introduce the new intrusion detection system named Intrusion Detection - Gain ratio Online Progressive Aggressive (ID-GOPA) model. This model uses offline machine learning in the first phase to train the model by the online Progressive Aggressive (PA) classifier to build a learnable model capable of being tested. The second phase is to start operating online using the trained model with the same prepossessing engine selecting only the relevant attribute based on information gain ratio algorithm and classifying every packet either as normal or attack in real-time detection. In addition we present and discuss simulation results, which confirm the superiority of proposed ID-GOPA compared to various attack detection algorithms such as SVM, NB, RF and DT.

Finally, we pointed out thesis achievement, limitations and future work in the conclusion and perspectives.

Chapter 1

Overview of Wireless Sensor Networks

1. Introduction

Communication systems are designed to transfer information from sender A to receiver B. One simple example is the mail service where sender A is composing a letter and taking it to the post office as illustrated in Figure 1.1. The mail is arranged and afterward conveyed through a vehicle to another post office. From that point, it is arranged and given to a mail transporter for delivery to the receiver. The letter may be dealt with at a few post offices and vehicles. Each piece of the mailing system is attempting to achieve exactly the same thing: conveying the mail to the destination. Postal services are built on a network of post offices, vehicles and working protocols which are contributing to deliver mail from A to B.

Per analogy, telecommunication network is defined as a set of components and wired links or wireless transmission channels that enable users to communicate and share data within same location or even across continents. Various types of networks have been emerging, starting by the smaller one like network on chips to the largest one such as deep space networks. A personal area network (PAN) can be used to connect user's laptop to its peripherals using Bluetooth technology. A wide area network relies on an undersea fiber optic cable to connect countries around the globe.

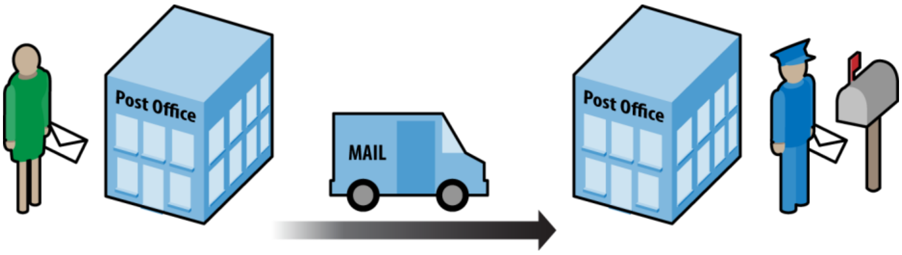


Figure 1.1 Postal system

In 1837, Samuel Morse invented the electric telegraph: an ingenious device using a sequential code with two elements: the line and the point. The first telegram between America and Europe was sent in 1866 via the transatlantic telegraph cable from Ireland to Newfoundland. The telegraph was limited to receiving and sending one message at a time. Alexander Graham Bell invented the telephone in 1876 while working on his own solutions to improve the telegraph via transmitting multiple messages over the same wire at the same time.

The evolution from telephony network and then data network continues over the last 150 years until full convergence of the telecommunication network and computer network into one network. Overall, many technologies contributed to this evolution leading to the development of internet as known today. Wired data network like Local Area Networks (LAN) and Synchronous Optical Networking (SONET) have born and growth in 1980s and 1990s. The early years of the 21 century have witnessed the development and expansion of Wireless technologies like Wavelength Division Multiplexing (WDM), which marked the emerging of high speed data exchange. Nowadays, new wireless standards continue to appear, supporting communication with cloud computing and data centers that are gradually becoming a foundation of today’s networking and computing world. Moreover, software defined network (SDN) is shaping networking function by separating control and data planes, and centralizing network intelligence leading to a real improvement of the Network efficiency and scalability [1].

2. Overview of network categories based on connection mode

At the present time, the development of wireless communication for mobile devices and web based applications, such as web browsing, online banking, online gaming and social media, has stimulated the wide spread usage of wireless network. Therefore, wireless networks have become a commodity and a vital component of contemporary daily life.

Advances in networking techniques have stimulated the emerging of a wide-range of network types. Networks can be grouped into two brands. The first class is wired networks, which rely on physical links such as wires and optical fibers. The second category is wireless networks, which use radio transmission techniques to establish links between nodes. Moreover, wireless networks can be split into two classes:

- Infrastructure based wireless networks that use fixed access points as gateways between wired and wireless area. For example, cellular networks (2G, 3G, LTE and 5G), Wi-Fi (IEEE 802.11), WiMax (IEEE 802.16).
- Infrastructures less networks are broadly known as Ad Hoc networks. This type of network does not rely on any pre-established infrastructure. In Ad Hoc networks, networking functions are distributed on all network components in order to create a self-organized, self-configured and self-administered system. Furthermore, Ad Hoc Networks may be single-hop like Bluetooth or multi-hop like Wireless Sensor Networks (WSN), Wireless Mesh Network (WMN) and Mobile Ad Hoc Network (MANET) [2].

3. Infrastructure based wireless networks

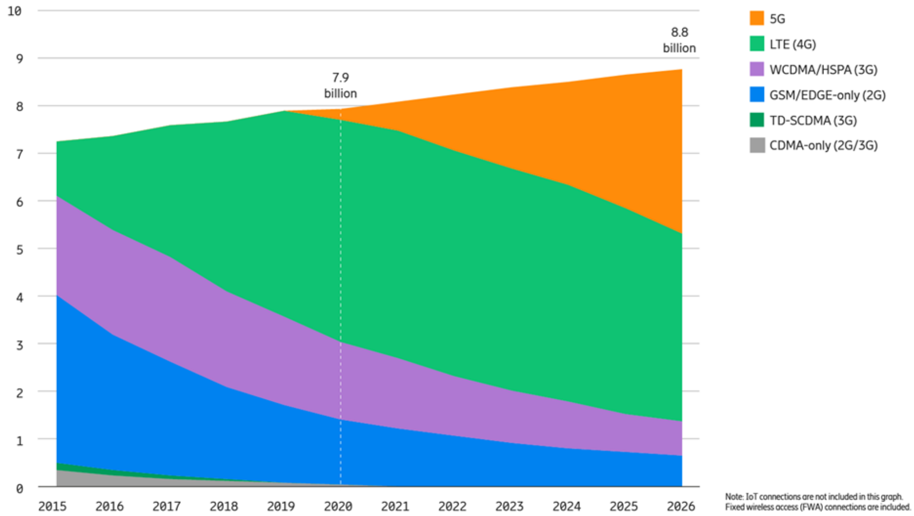
3.1 Mobile network evolution

Mobile networks have transformed the way individuals communicate and share knowledge. Advancement of wireless access technologies has reached its fifth generation (5G) since 2019. While a lot of the thrill around 5G technology focuses on its promise of increased speeds (from one to twenty Gigabit per second), it isn't only a quicker version of 4G. Its enhancements in reliability and latency will give varied edges, including internet access everywhere, real-time remote collaboration, precise location-sensing, device-to-device communication, and real-time visions. The wide success and the evolution of mobile services have been possible due to the development of end-user devices as well as the core network capabilities, figure 1.2 illustrates network evolution and associated services.

One of the 5G most significant characteristics is its enablement of ultra-reliable, ultra-responsive connections with latencies as low as one millisecond (ms). It takes 13ms for a picture seen by the human eye to be processed by the brain; which means with 1 ms latency, images transferred over 5G can be deemed as "instant."[1]. Ultra-low-latency 5G additionally offers the promise of "six nines," or 99.9999% reliability.[2]

[1] Anne Trafton, "In the blink of an eye," MIT News, January 16, 2014.

[2] Leonard Lee, "5G: The carrier-grade digital infrastructure for the software-defined factory of the future," CIO, September 9, 2019.



Note: IoT connections are not included in this graph.
Fixed wireless access (FWA) connections are included.

Figure 1.2 Mobile subscriptions by technology (billion) [3]

Back in the past, wireless access technologies have followed different evolutionary paths in order to guarantee high-level performance and efficiency in high mobile environment. The first generation (1G) was launched in 1982 using analog transmission to provide the basic voice mobile service; while the second generation (2G) were introduced in the beginning of 1990s. It improved system capacity and coverage area with the use of digital multiple access technology, such as time division multiple access (TDMA) and code division multiple access (CDMA). The third generation (3G) was introduced in the beginning of 21s by Generation Partnership Project (3GPP) organization. It delivered high-speed data service and open up the gates for mobile broadband reflection. The idea of mobile broadband was successfully realized in 2012 with the emerging of the fourth generation (4G). In addition to this advancement, the Fourth generation (4G) has offered access to wide range of mobile services, including data based mobile services and support of various mobility patterns required by applications in accordance with service demands in multi-user environment. The last generation, fifth generation, is expected to be more intelligent than its previous generations and achieves extra facilities more than offering worldwide interconnection [3]. In figure 1.3 we illustrate the timeline of mobile generation evolution.

[3] Source: Ericsson Mobility Report, November 2020. Mobile subscriptions outlook

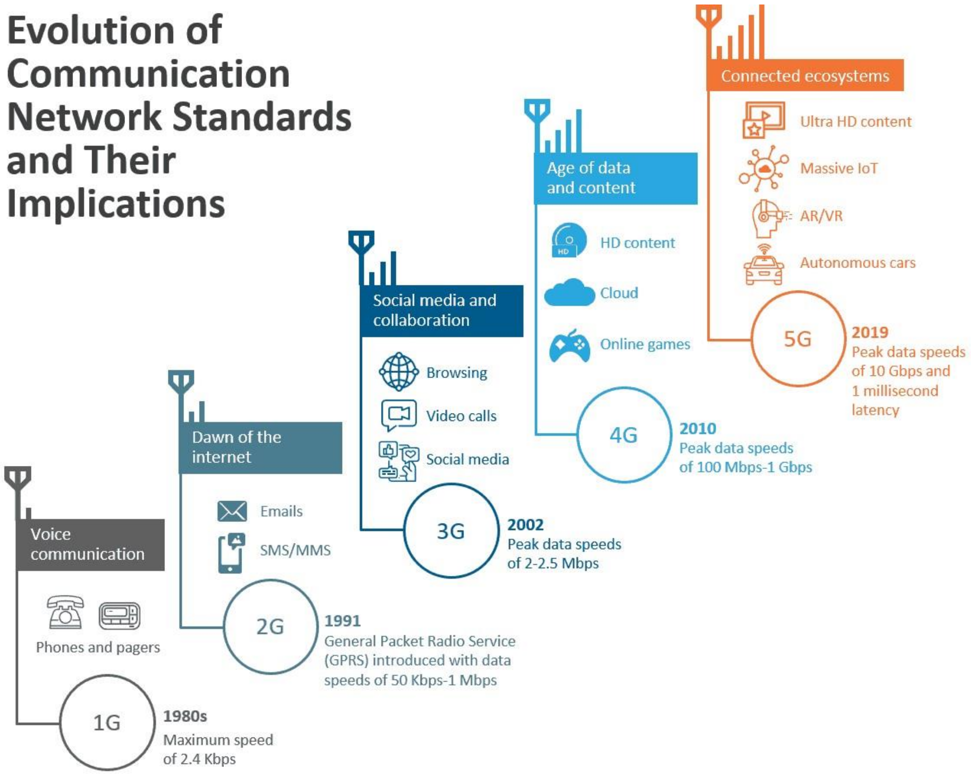

Evolution of Communication Network Standards and Their Implications
Connected ecosystems
Ultra HD content
Massive IoT
AR/VR
Autonomous cars
5G
2019
Peak data speeds of 10 Gbps and 1 millisecond latency
Age of data and content
HD content
Cloud
Online games
4G
2010
Peak data speeds of 100 Mbps-1 Gbps
Social media and collaboration
Browsing
Video calls
Social media
3G
2002
Peak data speeds of 2-2.5 Mbps
Dawn of the internet
Emails
SMS/MMS
2G
1991
General Packet Radio Service (GPRS) introduced with data speeds of 50 Kbps-1 Mbps
Voice communication
Phones and pagers
1G
1980s
Maximum speed of 2.4 Kbps

Figure 1.3 Timeline of mobile generation development[4]

With 5G features and capabilities, new applications which were until now impossible will be evolving in the coming years. These applications fall into 3 categories:

- **Enhanced mobile broadband (eMBB):** Allows connection speed enhancements between ten to twenty times quicker than today's 4G. It means 5G can stream high-definition video images from vehicles traveling as fast as 300 miles per hour without losing connection.[5]

- **Ultra-reliable low-latency communications (uRLLC)**: Offers high reliability and low latency, decreasing the time required to send a signal and receive a response to 10 milliseconds or less.[6] Low latency associated with high reliability open the door for mission-critical systems that don't tolerate any delay or error, including remote surgery and connected autonomous vehicles. Such applications will take time to be deployed until the 5G infrastructure is available in wide area. Figure 1.4 shows the

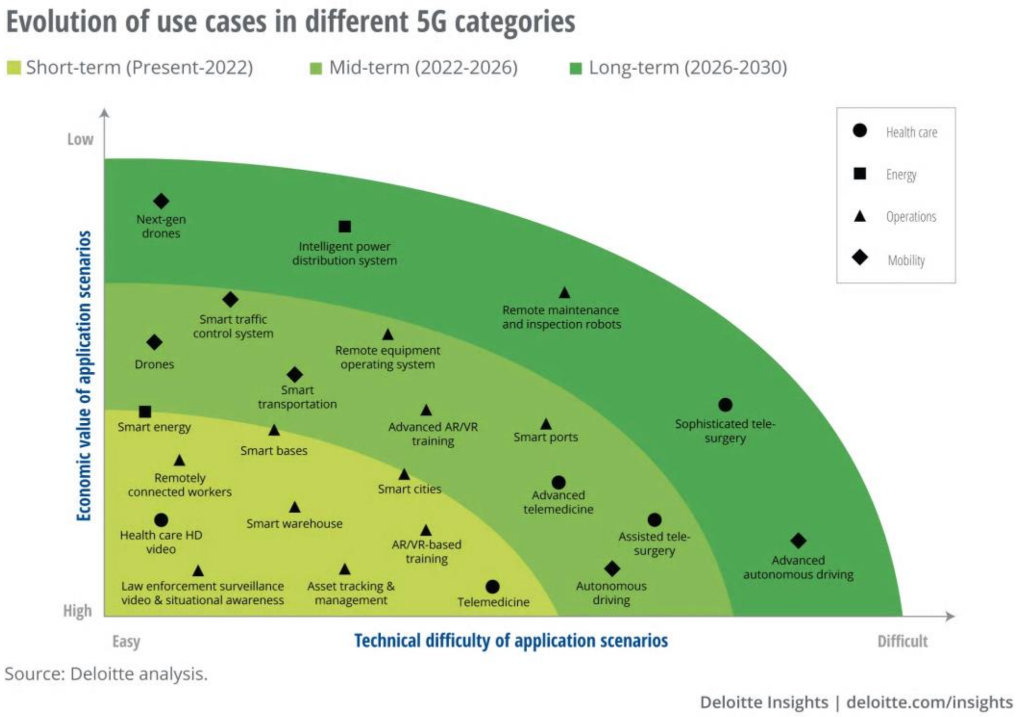
[4] Source: Evolution of Communication Network Standards and Their Implications | Market Insights™, December 18, 2019

[5] RF Wireless World, "Difference between 5G eMBB, mMTC, uRLLC."

[6] Bob O'Donnell, "5G latency improvements are still lagging," *Forbes*, February 18, 2020.

timeline of applications evolution in different 5G categories depending on their deployment difficulty and their economic value.

- **Massive machine-type communication (mMTC):** Enables connecting up to 1 million device per square kilometer, versus 100,000 maximum for 4G. This high density will allow to connect everything in the future.[7]



Source: Deloitte analysis.

Deloitte Insights | deloitte.com/insights

Figure 1.4 Evolution Timeline of the 5G applications scenarios [8]

3.2 IEEE 802.11 (Wi-Fi)

Wi-Fi or the IEEE standard 802.11 is a common wireless technology used at home or workplace by small to large businesses and service providers in order to establish a private wireless local area network with ranges of about hundred meter between end-user device and the Access point. As shown in figure 1.5, Wi-Fi and LAN can be combined in a network offering both connectivity options. There are two types of Wi-Fi devices:

[7] Paul Lee, Mark Casey, and Craig Wigginton, *Private 5G networks: Enterprise untethered*, Deloitte Insights, December 9, 2020.

[8] Source: 5G in government: The future of hyper connected public services, Deloitte Insights, August, 2020.

- Standard devices are designed for end users and widely available at competitive price in computer stores;
- Enhanced Wi-Fi devices are designed for ISP use;

The IEEE committee introduced the first release of IEEE 802.11 standard in 1997. However, this release did not receive big success because of its relative low data rate and relative high cost. Therefore, further standardized products (such as 802.11b, 802.11a, 802.11g, and 802.11n) were introduced with enhanced features and were much more successful.

Here after some advantages of Wi-Fi:

- As a standard it's universal and vendor neutral; all Wi-Fi devices are compatible with each other regardless of the constructor;
- Affordable cost;
- Ability to extend the range and performance of a Wi-Fi network; Scalability of the network.

Disadvantages of Wi-Fi are as follows:

- It is not suitable for wide area networking (WAN); Limited distance of coverage.
- It uses carrier sense multiple access (CSMA) mechanism. Only one wireless node can send data at a time within same access point, which means in the case of one access point, one user can dominate all network's resources. Consequently, real time applications such as video conferencing, voice over internet protocol (VOIP), and multimedia can break down the network [4].

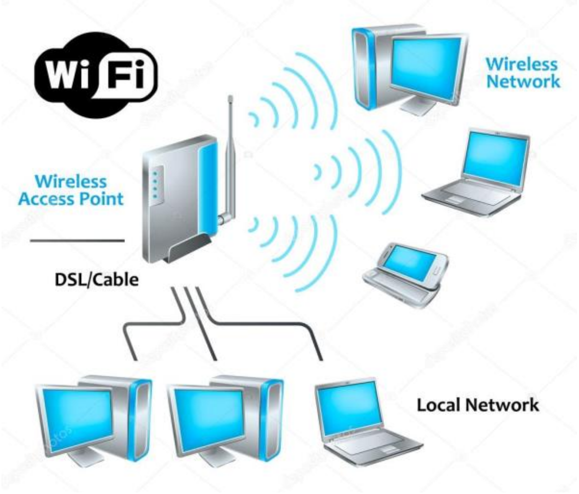


Figure 1.5 WiFi topology[9]

[9] Source : https://fr.depositphotos.com

3.3 IEEE 802.16 (WiMax)

WiMax or 802.16 is a standard-based technology enabling the delivery of last mile wireless broadband access as an alternative to the digital subscriber line (DSL) system, as illustrated in figure 1.6. It is an attractive alternative to network operators to offer broadband access service in geographically remote area with no or limited wired network. WiMax promises high-speed data rate, initially offered about 30 to 40 Mb/s then updated to 1 GB/s in 2011 and offered extended coverage area. It is suitable to establish metropolitan area networks (MAN) or wide area network (WAN) [4][5].

The advantages of WiMax technology are as follows:

- Specially designed for MAN and WAN;
- Saving in speed of deployment;
- Installation cost effective;

Disadvantages of WiMax are the following:

- New technology that has not passed the test of time;
- More expensive than Wi-Fi;

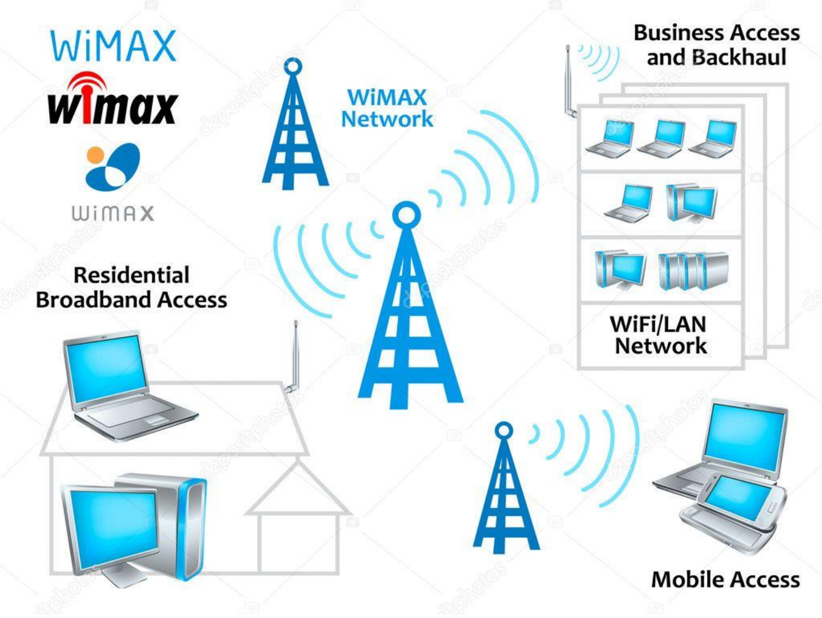


Figure 1.6 Generic topology of WiMax network[10]

[10] Source: https://fr.depositphotos.com

4. Infrastructure less networks

The increased need to develop flexible and cost effective telecommunication network has made wireless networks a promising and popular field. Therefore, infrastructures less wireless networks have become hot topics of high interest in research fields. It is distinguished by a decentralized communication scheme, infrastructure-less environment, easy deployment and maintenance. These distinguishing features are well suited for military operations and disaster management [6]. There are two types of infrastructure less network, Single-hop network and multi-hop network, depending on how many nodes are involved in data forwarding session.

4.1 Single-hop network

4.1.1 IEEE 802.15 (Bluetooth)

Bluetooth is an open wireless technology standard for short-range radio frequency communication. Bluetooth technology, offers peer-to-peer communication as illustrated in figure 1.7. It was primarily used to establish wireless personal area networks (WPANs), and it was integrated into many types of business and consumer devices. Ericsson created Bluetooth in 1994 as a wireless alternative to RS-232 data cables. It can connect several devices and overcomes synchronization problems. Furthermore, Bluetooth is a high-speed, low-power microwave wireless link technology designed to connect easily phones, laptops, personal digital assistants (PDAs), and other portable devices. Unlike infrared communication system, Bluetooth does not require line-of-sight positioning of connected devices. It uses some adaptations of current WLAN techniques and it is most distinguished for its small size and low cost. Whenever any Bluetooth-enabled devices get within the range of each other, they immediately transfer address information and establish small network between each other, without the need of user assistance. Data can be transferred between master and slave devices where the master chooses which slave device to address. It rapidly switches from one device to another in round-robin fashion [4][7].

Figure 1.7 Bluetooth peer-to-peer connection[11]

The main features of Bluetooth technology include the following:

- Operates in the 2.56 GHZ band, which is globally available free of charges;
- Uses frequency hop spread spectrum (FHSS);
- Can support up to eight devices in a small network;
- Omnidirectional, don't require line-of-sight transmission;
- Up to 100 m range;
- Low cost.

4.2 Multi-hop network

Since last decades, we have been assisting to the development of shared applications that require networking facilities to deliver virtual online services anywhere at any time, such as remote storing and processing in cloud service platforms and remote sensing in internet of things applications. In this context, Ad hoc networks have been massively used in wide applications' areas, particularly in controlling and monitoring environment parameters, weapon control and tracking, urban intelligent transportation, smart cities and other fields [8].

The concept of ad hoc networking concept has broad field of specific applications which opened up the opportunity to develop multiple network sub-categories like Mobile Ad Hoc Network (MANET) [2], Vehicular Ad Hoc Network (VANET) [9], Flaying Ad Hoc network (FANET) [10] and Underwater Ad Hoc Network (UANET) [11].

4.2.1 MANETs

Mobile Ad Hoc network is an autonomous system formed by collection of intelligent mobile nodes. These nodes rely on wireless communication techniques to establish temporary

[11] Source: https://fr.depositphotos.com

wireless links in order to accomplish data transfer between a pair of source node and destination node. As shown in figure 1.8, MANET is infrastructure less based wireless network, where nodes do not rely on any pre-existing infrastructure to establish connections between themselves. Thus mobile nodes constitute a self-organized, self-configured and self-administered system. Due to limited coverage range of mobile nodes, most of time communication is established through several hops. This operation is known as multi-hop networking where source node might need the contribution of intermediate nodes to deliver data flows to destination node.

The random and unpredictable movements of nodes in MANET cause fast and unexpected network topology changes. Routing path in MANET potentially contains multiple-hops, and every node in the network has to play both roles router and host [12].

Since Ad Hoc wireless network is different from the classical wireless and wired networks, a distinctive set of challenges are present while implementing such networks. Routing protocol in Ad Hoc network is considered as one of the prominent challenges.

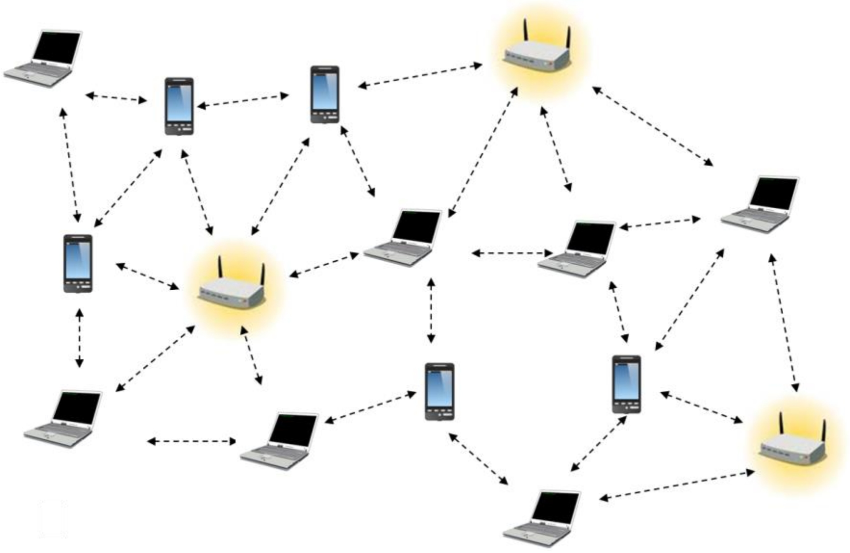


Figure 1.8 Mobile Ad Hoc network[12]

4.2.1.1 VANETs

Vehicular ad hoc networks (VANETs) are special type of mobile ad hoc networks (MANETs) where wireless-equipped vehicles form spontaneously the network while traveling along the road. Direct wireless transmission from vehicle to vehicle makes communication

[12] Source: How to create adhoc network, by shivam srivastava, 2015

feasible even if telecommunication infrastructures are not available, such as mobile base stations or access points of wireless dedicated access networks. In recent years, this new way of communication has been attracting much attention in academic and industry communities. The US Federal Communications Commission (FCC) has allocated seven 10-MHz channels in the 5.9-GHz band for dedicated short-range communication (DSRC) to enhance safety and productivity of transportation system. The FCC's DSRC ruling has permitted both safety and non-safety (commercial) applications, provided safety is assigned priority. The IEEE established new standard for VANETs called IEEE 802.11p [13].

Communication pattern in VANETs includes two forms: Vehicle to vehicle (V2V) communications and vehicle to infrastructure (V2I) communications (figure 1.9). The former V2V leads to a pure MANETs, while the latter V2I can be seen as hybrid network. Although VANETs can be seen as special case of MANETs, there are several distinctive characteristics that involve special treatment of VANETs [14]. The most important distinctive characteristics of VANETs are as follows:

- Specific mobility patterns
- Highly dynamic topology
- Intermittent connectivity
- Strict quality of service (QoS) requirements

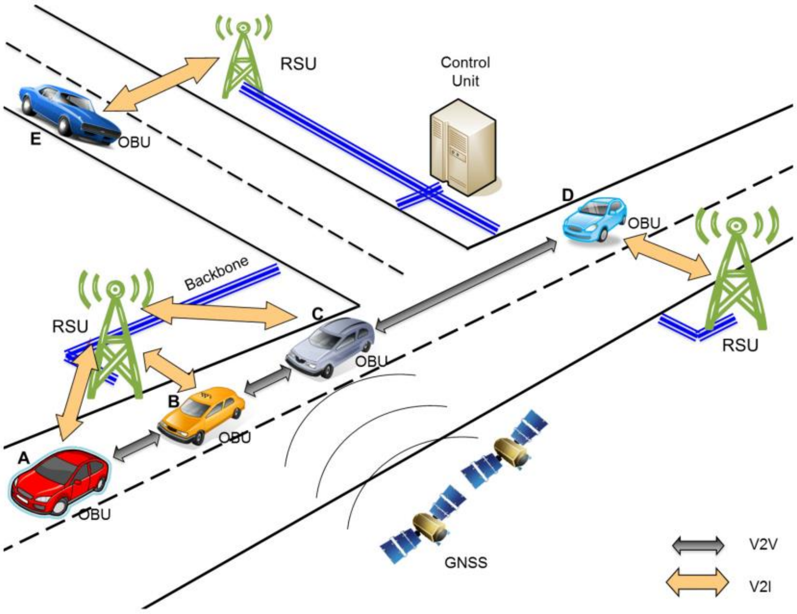


Figure 1.9 V2V and V2I communication[13]

[13] Source: Sensing Traffic Density Combining V2V and V2I Wireless Communications. Sensors 2015

4.2.1.2 FANETs

Flying Ad-Hoc Network (FANET) is a collection of small-unmanned aerial vehicles (UAVs). It represents a special case of mobile ad hoc networks (MANETs). The UAVs fly in the sky and communicate through each other with the help of satellite or ground base station to establish an ad-hoc network as illustrated in figure 1.10. This makes them one of attractive technologies for many civilian and militaries applications [10]. FANET's topology may change more frequently compared to MANET and VANET due to high mobility of UAVs. This make system design pretty challenging and require cooperation and collaboration between UAVs [15].

Ad hoc network between UAVs is one of the most effective communication architectures for multi UAV systems. Although, UAV cannot communicate directly with the base station or satellite, it relies on multi-hop communication scheme, to guarantee that all UAVs are connected to each other and to the base station or satellite all the time without any infrastructure. Furthermore, most UAV execute real time operation, where high data rate is required. This leads to high bandwidth requirement compared to MANET or VANET. In addition, FANET require high speed compared to MANET and VANET. Therefore, high gain antenna is required to achieve longer range, reduce hop count and enhance the overall performances [16].

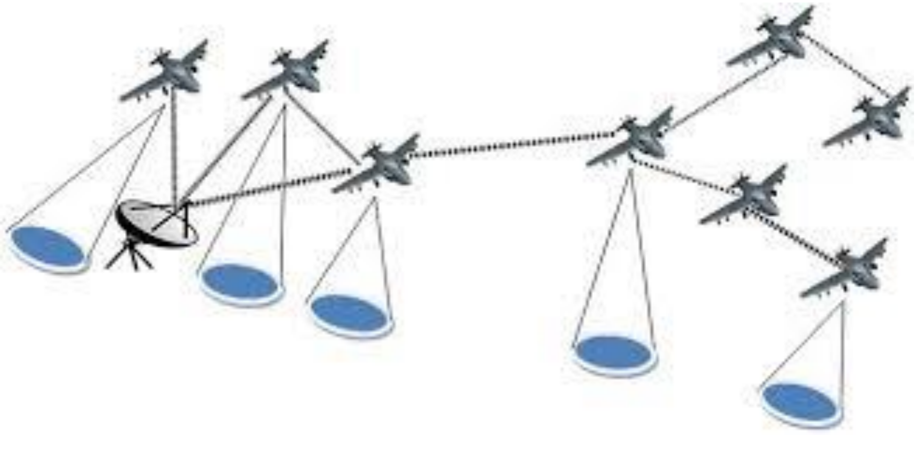


Figure 1.10 Fanet network[14]

[14] Source: A Survey: Different Mobility Model for FANET

4.2.1.3 UANETs

In recent years, underwater ad hoc networks (UANETs), illustrated in figure 1.11, have become an important field of research due to its important applications such as oceanographic data collection, ocean sampling, environmental and pollution monitoring.

Underwater communication system is fundamentally based on three techniques for transmitting data. These methods are based on electromagnetic (EM) waves, optical signals and sound waves. Each one of these techniques has advantages and disadvantages that depend essentially on physical constraints.

Firstly, the use of EM waves to transmit signals in the water is characterized to be a fast and efficient communication between network nodes. Secondly, the use of optical signals are generally limited to short distances because water has a fairly high absorption factor in optical waveband and lastly the use of sound waves in water is more much efficient than in air because in water sound waves propagates faster and has lower energy losses than in air. For example, sound waves are transmitted into the sea at a speed range of 1400 to 1600 m/s, while speed propagation is 340 m/s in air.

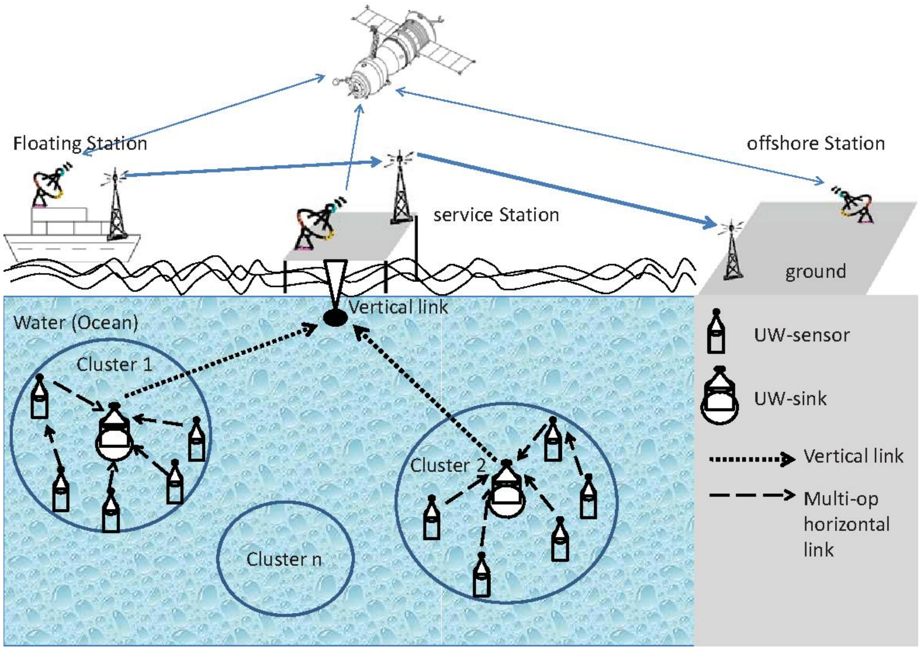


Figure 1.11 Underwater wireless Ad hoc network[15]

The main differences between underwater and terrestrial ad hoc networks are as follows:

[15] Source: Underwater acoustic sensor networks: research challenges, Elsevier, 2005

- Cost: Underwater devices are expensive;
- Density: Underwater deployment are generally sparser than terrestrial ones;
- Power: Acoustic underwater communication requires more power than terrestrial radio communication;

4.2.2 Wireless Sensor Networks (WSN)

Sensor network consists of large number of very small nodes that are deployed in some geographical area. The main purpose is to sense the environment and report what is happening in monitored area. Sensor network is used in many applications such as surveillance and target tracking in military field. In industrial applications, sensor network helps monitoring hazardous chemicals. They are also helpful in environment monitoring and early fire warning in forest as well as seismic.

Figure 1.12 shows the typical configurations of wireless sensor networks. Nodes are scattered in a geographical area, which is divided into clusters. Nodes in each cluster communicate with the routing node that aggregate data and send it across a gateway to the user.

Typically, sensor network works in one of two modes: Continuous or query mode. In continuous operation mode, the node is continuously sensing the environment and sending data to neighboring or central node. In query mode, the node is usually powered down waiting command from central node, or neighboring node. When node receives the commands it collects data from sensor, processes it and sends it to requesting node [17].

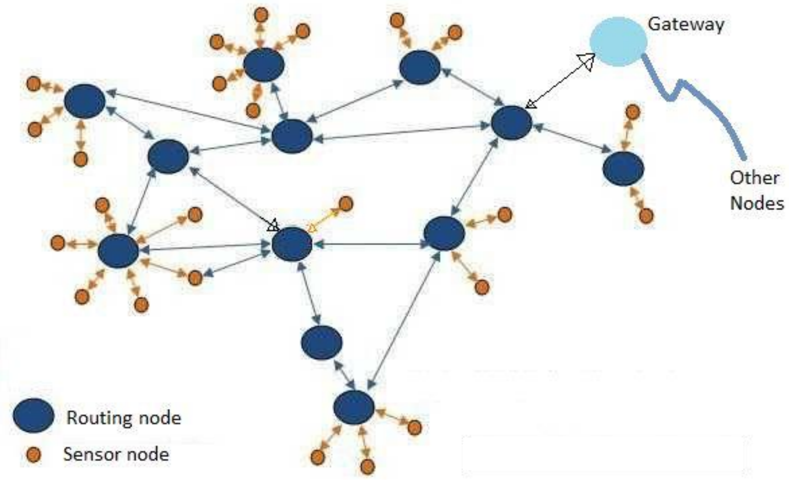


Figure 1.12 WSN network[16]

[16] Source: Brilliant Sensor Network Architecture, http://cialisalto.com/simple-sensor-network-architecture/

4.2.3 Wireless Mesh Networks (WMN)

Wireless mesh networks (WMN) have emerged as vital technology for various applications such as broadband home networking, districts networking, enterprise networking and metropolitan area networking. As illustrated in Figure 1.13, typical topology of WMN is composed of three distinct wireless network components. The fist one is mesh gateway that act as mesh routers with gateway/bridge futures. The second one is mesh routers or access points and the last one is mesh client. Mesh clients connect to mesh routers using wireless or wired links. Every mesh router achieves relaying data for other mesh routers, and some mesh routers support extra capabilities to act as internet gateways. Generally, router gateway often has wired link to transfer data traffic between mesh routers and internet network [18].

The WMN has attractive advantages such as: self-organization, self-healing, self-configuration, enabling quick deployment, easy maintenance, and cost effectiveness. WMNs inherit almost all characteristics of wireless ad hoc networks like decentralized design, distributed communications. However, unlike the mobility of ad hoc nodes, mesh routers are usually fixed.

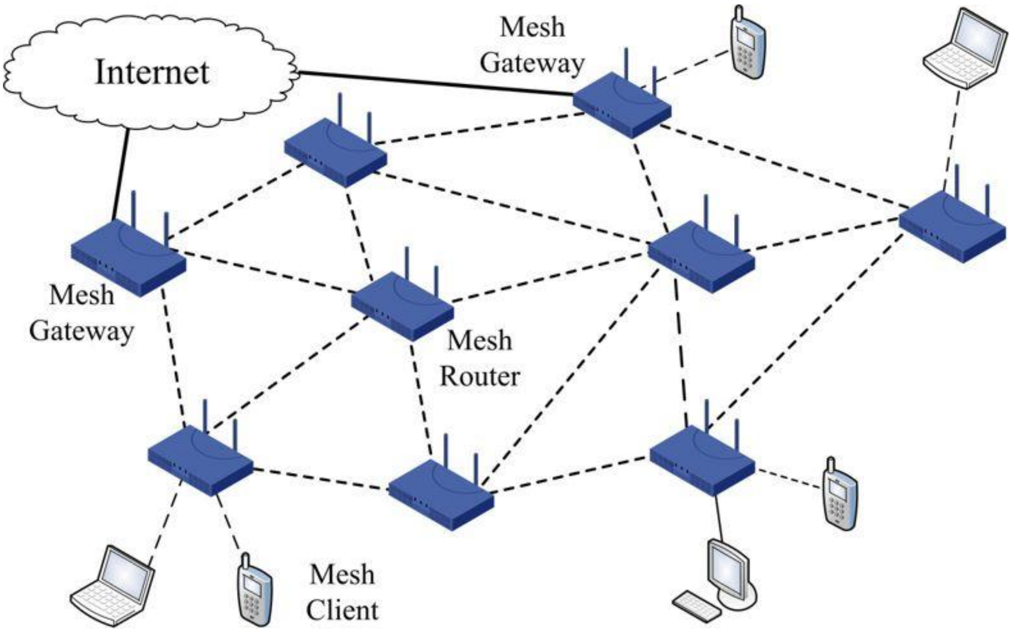


Figure 1.13 WMN network topology[17]

[17] Source: F. Liu and Y. Bai, “An overview of topology control mechanisms in multi-radio multi-channel wireless mesh networks,”

5. Evolution of Wireless Sensor Network

The concrete illustration of a wireless network that uses the endpoints as a core network like in the contemporary WSN started with the Sound Surveillance System (SOSUS), developed by the United States Military within the1950s to notice and track Soviet submarines. The network is based on distributed hydrophones, which are acoustic sensors submerged within the Atlantic and Pacific oceans. This sensing technology remains operational nowadays, albeit serving a lot of peaceful functions of observance oceanic life and volcanic activity.

The second generation were called Distributed Sensor Network (DSN) and emerged via a program launched in the1980 by the United States Defense Advanced Research Projects Agency (DARPA). DSN explored the challenges in implementing distributed/wireless detector networks. With the birth of DSN and its progression into academia through partnering universities like UC Berkeley and the Massachusetts Institute of Technology Lincoln Labs, WSN technology found a way to attract civilian research projects.

In 2001- 2002, many startups emerged around sensor networks industry like Sensoria, Crossbow and Ember Corp. The goal of those initiatives is to enable high-volume of WSNs in the light of wide industrial usage via reducing the price and energy per device, whereas simplifying development and maintenance tasks.

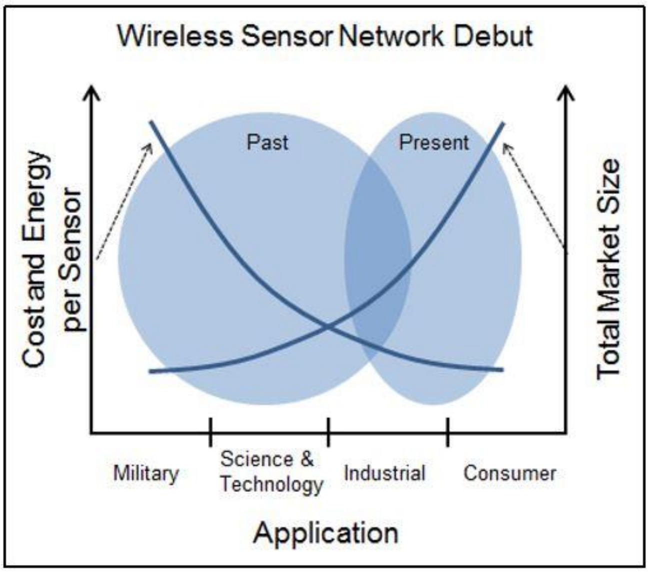


Figure 1.14 WSNs Gain Market Traction with Decrease in Sensor Costs[18]

[18] Source: The Evolution of Wireless Sensor Networks, Silicon Laboratories, Inc. Rev 1.0

Wireless Sensor Network is an infrastructure less mobile network, commonly known as the WSN, which is a self-organizing and self-configuring multi-hop system. It does not require any fixed infrastructure. In such network, nodes are dynamically and arbitrarily located, and are required to relay packets for other nodes in order to deliver data across the network [19].

6. IEEE Standard 802.15.4

Wireless Sensor Network standard of communication has been defined in the IEEE 802.15.4. The purpose of the IEEE P802.15.4 is to produce a standard that enabled very low-cost and low-power communications. One of fields of applications of this technology is the implementation and execution of WSNs.

The last revision of the standard was published in July 2020 to include six approved amendments subsequent to the 2015 revision. This revision added 2 physical layers (PHYs) amendments and one MAC amendment, with corrigenda and clarifications. The features added by the amendments include the following:

— A variety of new PHY modulation, coding, and band options to support a wide variety of application needs including smart utility networks (SUNs), ternary amplitude shift keying (TASK), china medical band (CMB) and rate switch Gaussian frequency shift keying (RS-GFSK)

Most of the changes and clarifications were following requests from individuals after the revision in 2015. Major corrigenda items included changes to the transmission order of the address field.

The first standard, IEEE Std 802.15.4-2003, defined 2 physical layers (PHYs), operating in different frequency bands with a simple and effective medium access control (MAC).

In 2006, 2 optional physical layers have been added to the revised standard. The MAC compatibility with previous version was not touched, but the revision added MAC frames with an increased version number and a variety of MAC enhancements, including the following:

— Introduce data time stamping mechanism to manage shared time base

— Support for beacon scheduling

— Synchronize broadcast messages in beacon-enabled personal area networks (PANs)

— Improved MAC layer security

In 2011, revision of the standard was created to include the 3 amendments approved subsequent to the 2006 revision. This revision added 4 physical layers options alongside with the MAC capability to support ranging. Furthermore, the organization of the standard was changed so that each PHY would have a separate clause, and the MAC clause was divided into 3 sub clauses: functional description, interface specification, and security specification.

In 2015, a revision of the standard was created to roll in the amendments approved subsequent to the 2011 revision: 6 PHY amendments and one MAC amendment, with corrigenda and clarifications. The features added by the amendments include the following:

— Enhanced frame formats maintaining backward compatibility

— Information Elements (IEs)

— Channel agility

— Extended super frame options

— Low-energy mechanisms

— Enhanced acknowledgment setting that can carry data and can be secured

— Prioritized channel access

— A selection of new physical layers modulation, coding, and band options to support a variety of application needs including radio frequency identification (RFID), smart utility networks (SUNs), television white space (TVWS) operation, low-energy critical infrastructure monitoring (LECIM), and rail communications and control (RCC)

By the late 1990s, the semiconductor industry started a phase of standardizing the technology for making low power integrated circuits. This technology named Complementary Metal Oxide Semiconductor (CMOS) reached an attractive cost points for large sensor network deployments. The following hardware improvement was the system-on-chip (SoC) which include enough high-performance peripherals like amplifiers, analog-to-digital converters (ADCs), and storage to handle both the application processing and network protocol stack while communicating with the network nodes via the Radio-frequency (RF) transceiver.

WSN is an infrastructure less network which is usually a self-organizing and self-configuring multi-hop system. It does not require any fixed infrastructure. In such network, nodes are dynamically and arbitrarily located, and are required to relay packets for other nodes in order to deliver data across the network [19].

Even with all promises that are offered by wireless sensor networks, successful commercial deployment requires realistic solutions to different problems, including data protection and real-time applications, pricing, distributed collaboration, energy-efficient management, data aggregation, and support for multicast traffic [20].

7. Characteristics of WSNs

Wireless Sensor Network is an autonomous network where each node auto-discover its neighbors. All nodes participate to create the network map and routing tables. These tiny devices have limited power (generally a battery for the full sensor lifetime) and reduced computing resources and memory. WSN are deployed for many applications like environment monitoring and health monitoring. In order to reduce the energy consumption and increase the WSN life, adapted network protocols like routing and data aggregation are used. The scheduling of the communications also allow better management of active/idle status and hence optimize the energy consumption.

The main features that characterize WSNs, from others infrastructure based networks, are the absence of an intelligent central platform which is generally responsible of controlling and managing edge nodes; like base station in cellular networks. They are in charge of connecting cellular phone of end users in order to operate voice calls or data connections to the internet or corporate intranet. Thus in WSNs, intelligent processes are equally shared by sensor nodes, which are involved in an active communication at a specific time. Each sensor node keeps sensing their neighboring to acquire and maintains network topology that might change due to mobility of nodes or nodes disconnection. Moreover, there are some limitations inherent to wireless communication technologies such as limited bandwidth, variable links capacity, limited physical security, multi-hop routing and others features inherent to harsh environmental conditions where nodes are deployed such as fire area or battlefield, network topology, device heterogeneity and limited energy [21] that makes WSNs operations complex and challenging.

Figure1.15 shows a typical architecture of sensor node used in WSN. Sensor nodes are responsible for gathering the physical information of environment before transforming it into a digital signal via Analog to Digital Converter (ADC) and sending it to a sink node. The WSN is self-organized to manage communications and routing between nodes until reaching the base station (BS).

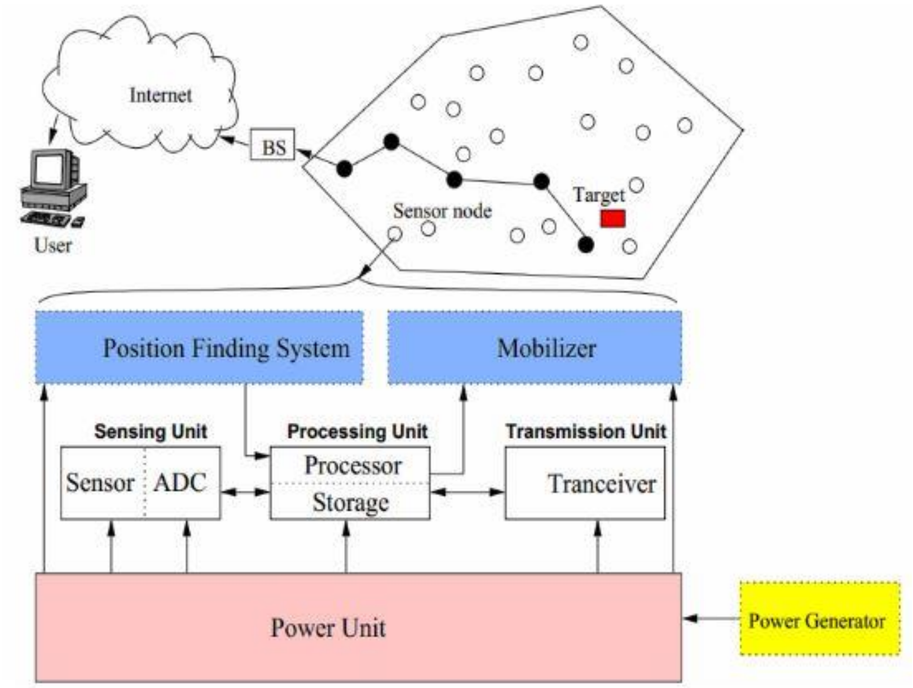


Figure 1.15 Sensor Node Architecture[19]

8. Applications of WSNs

With the development of very small and cheap sensors, expansion of mobile devices market and rapid progress in wireless communication; wireless sensor networks is gaining importance with the widespread of related applications. In this context, wireless sensor networks offer numerous opportunities to develop physical world monitoring applications and networking facilities in the areas where fixed infrastructure is not available, too expensive to deploy or there is no time to set up it [20].

Wireless sensor network allows to monitor temperature, humidity, vibrations, light or gas as well as many other physical properties in harsh environment while easily adding and removing devices to and from the network. Diverse sets of WSN applications are available ranging from large-scale high density networks to small networks with the common constraint of power shortage as every sensor is using a battery as source of energy for all its lifetime. A

[19] Source: https://www.researchgate.net/publication/331928309_MANET_and_WSN_WHAT_MAKES_THEM_DIFFERENT

great deal of new services is evolving as sensing additional physical properties becomes possible with high accuracy. Typical applications include:

- **Military Applications**:

Wireless Sensor Network can be providing precious information's for military intelligence via monitoring the movement of the enemy. Battlefield Surveillance uses sensor nodes deployed on a battlefield nearby of the routes that enemy vehicles and soldiers may use. WSN can be deployed quickly and start delivering information's like acoustic signals produced by moving target objects[22]. This network which doesn't require any maintenance can detect and categorize various objectives, such as vehicles and troop movements, based on the spatial differences of the signal strength detected by the sensors [23], the location of the enemy forces can be estimated.

Intruder detection is also possible based on acoustic and seismic sensors monitoring the sounds and vibrations of soldiers and weapon actions. Intruder's detection is mainly useful in difficult terrains where visibility is limited because of heavy rain or large vegetation for example[24].

Figure 1.16 shows how sensors can monitor the battlefield and the forces movements while transferring real time data across the WSN to either a static or mobile sink which could be a computer on the ground (carried by a soldier or a vehicle) or in the sky (drone, helicopter..etc).

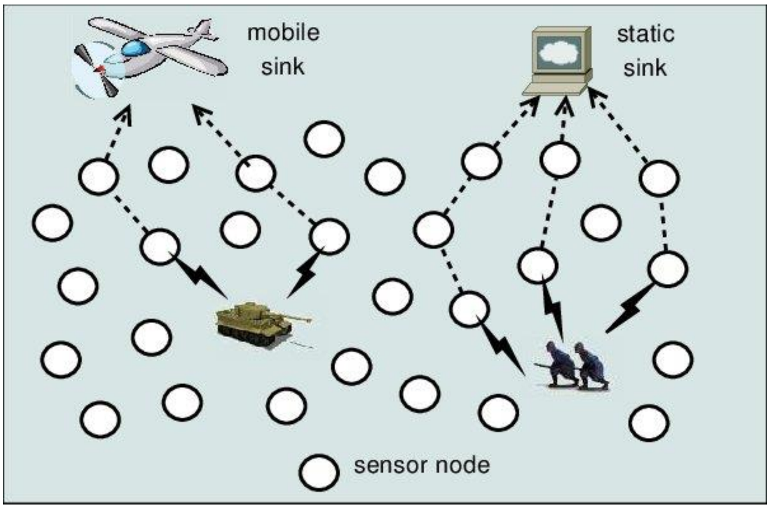


Figure 1.16 WSN Combat and Battlefield Monitoring[20]

- [20] Source: Strategies for Data Dissemination to Mobile Sinks in Wireless Sensor Networks, IEEE Wireless Communications 15(6):31 – 37, January 2009

- **Medical/Health Applications:**

Wireless sensor networks can use biometric sensors to monitor patient's health either within the medical facility or remotely. In [25] authors presented a home-based wireless ElectroCardioGram (ECG) monitoring system using Zigbee technology. The collected data can be stored in the patient mobile and viewed either real time or periodically by the physicians. This allows patient to live in their home while getting medical services remotely. Health monitoring system can monitor simultaneously multiple physiological signals providing important and complete data to the physicians. Another advantage of remote health monitoring is reducing dramatically the costs of having a bed in the hospital. In India, giving a lot of clinical system at home will enable passing on pretty much 70% of each clinical organization at home[26].

Figure 1.17 illustrate a practical case of remote monitoring of multiple physiological signals (heart rate, respiratory rate and motion). Data is consolidated at the mobile device which is playing the Sink role and from there the bridge within internet allows communication with medical health givers. In case of anomaly detection, care givers can implement intervention.

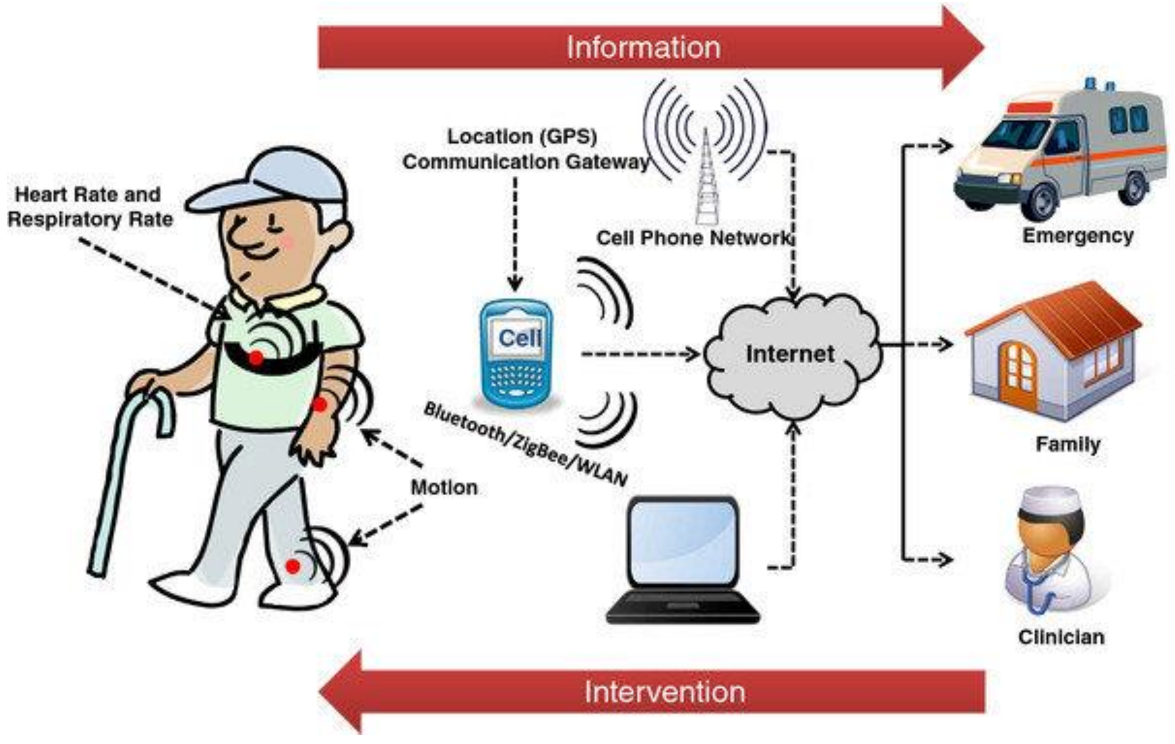


Figure 1.17 Remote health monitoring system based on wearable sensors.[21]

[21] Source: A Review of Wearable Sensors and Systems with Application in Rehabilitation, April 2012

- **Environmental Applications:**

It includes air monitoring to measure the pollution and the effect of human activities on the air. Sensors measure the air quality parameters based in gazes such as CO and NO2. Water monitoring cover both fresh drinking water and oceanic water[27]. Fish farm water monitoring help detecting water quality degradation and preventing higher damage to the flora and fauna. This alert allows to trigger another analysis of the source of the pollution which could be from feed and fecal waste.

Volcanic Activity Monitoring can save lives if the alert is sent on time to people close to the danger before the eruption. WSN can detects some volcanic signs which are indicating a soon eruption.

Forest fire detection and prevention: The effect of a forest fire can be devastating to all kinds of life and to the economy. In 2019-2020 the bushfire in Australia burned 18.6 million Hectares leading to extinction of some species with about three billion terrestrial vertebrates affected, the vast majority being reptiles. 34 people lost their life and almost 5900 building destroyed. The economic cost exceeds $100 billion making fires the most costly Australian natural disasters[22].

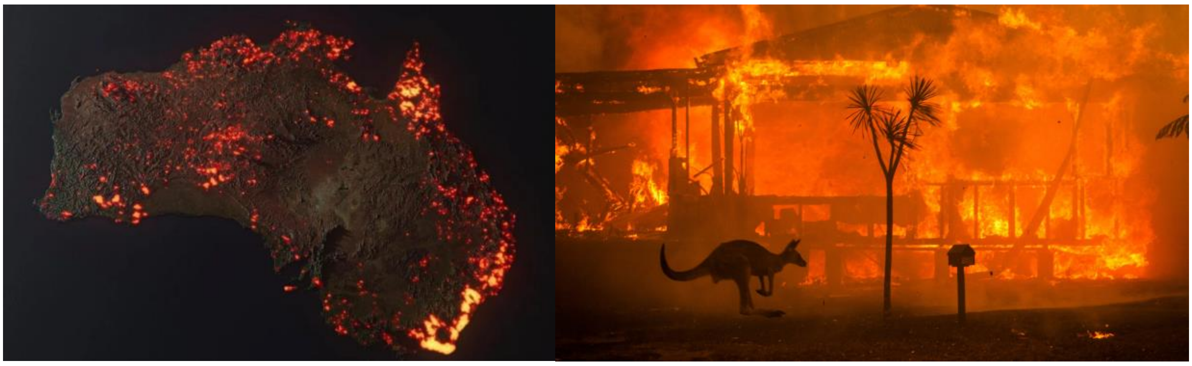


Figure 1.18 2019-2020 Fire in Australia.[23] [24]

The image to the left in Figure 1.18 is not a photo but a 3D visualization of the fires in Australia. Photograph Anthony Hearsey, made it based on data from NASA's FIRMS (Fire Information for Resource Management System). The 3D visualization is consolidating the areas which have been affected by bushfires between 05/12/2019 – 05/01/2020 from NASA's

[22] Source: 2019–20 Australian bushfire season, Wikipedia, last update May 2021

[23] Source: Isabella Kwai, New York Times, Published Dec. 31, 2019Updated Jan. 23, 2020

[24] Source: David Mikkelson, Is This a Photo of Australian Fires Taken from the Space Station? , January 2020, https://www.snopes.com/fact-check/australia-fires-iss-image

Satellite data gathering regarding fires (FIRMS). The photo to the right shows a kangaroo rushes past a burning house during Australia big fire. The New York Times describe it as one of “Apocalyptic Scenes in Australia as Fires Turn Skies Blood Red”.

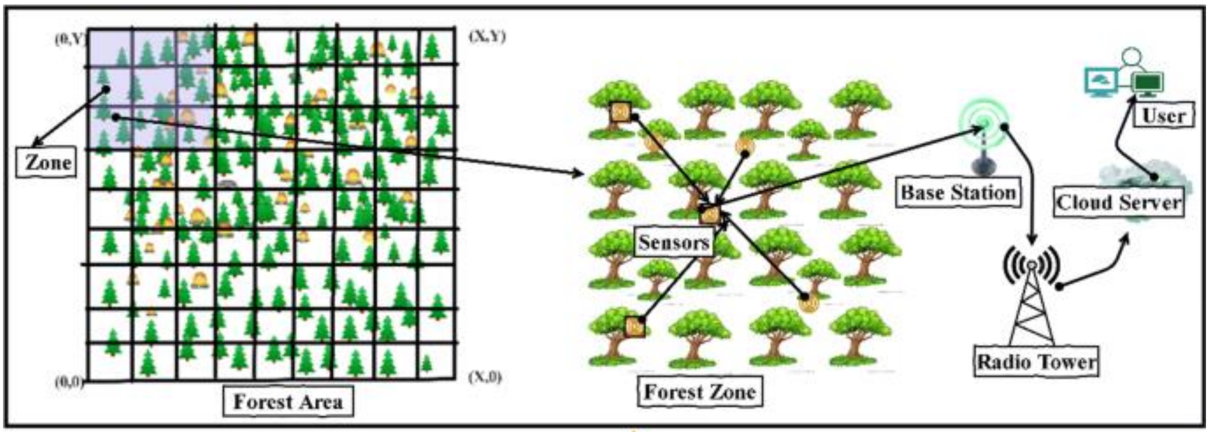


Figure 1.19 Forest Fire monitoring using WSN[25]

Figure 1.19 shows a fire monitoring system using Wireless Sensor Network [28]. The sensors are monitoring temperature and humidity to expect high risk zones and follow the fire when happening to guide firemen during their interventions[29].

In [30], the authors made some remarks under normal conditions related to temperature and humidity values showing a recurring changes during the different stages of the day where temperature and humidity values vary very slowly in opposite ways. One sign of fire would be then the higher rate of variation of temperature and humidity. The study excludes the effect of Sun rays on the sensors as the temperature shows a rate of change similar to that produced by a fire.

Seismic Activity Monitoring [31] and Tsunami monitoring are also based on vibration monitoring and can save lives when informing early of such catastrophe. Tsunami detection uses WSN with sensors deployed underwater in coastal regions.

- **Agriculture & Crop Monitoring:**

WSN helps the farmers to monitor temperature, light intensity, humidity, irrigation system, measuring water supply and so on [32]. These are key factors for controlling the crops quality and increasing the produced quantity and reduce the cost of yield.

- **Robotics & Industrial Applications:**

[25] Source: Raj Vikram, Ditipriya Sinha, Debashis De & Ayan Kumar Das; EEFFL: energy efficient data forwarding for forest fire detection using localization technique in wireless sensor network, Springer, 16 June 2020

Robotics and WSN have been attracting a lot of attention from research perspective but putting the two fields into collaboration is still offering opportunities for managing some challenges. The authors in [33] highlighted the core problems and research trends related to Robotic Wireless Sensor Network (RWSN) such as connectivity, localization, routing, and robust flow of information.

RWSN allow robots to communicate real time and exchange data like their own position or the manipulated objects as well as monitoring other parameters like pressure, temperature or vibration. Nowadays, the usage of industrial wireless sensor networks is increasing in several industries such as healthcare, automotive, gas and oil and other manufacturing sectors. Low costs and quick and easy deployment of wireless sensor networks is generating a high demand for enhanced technology, processing efficiencies [34]. It also reduce the operators risk exposure in some environments. The combination of WSN technologies, Robotics and the human expertise are promising leading industries high revenue in the global industrial wireless sensor networks market.

9. Challenges of WSNs

Regardless of the attractive applications discussed in section 7, WSNs features introduce several challenges that must be carefully reviewed before wide commercial deployment can be expected. On top of sensing functionality, sensor nodes manage tasks related to network routing, data protection or challenges are to be faced by the sensor nodes with this additional responsibility

In WSNs environment the task of connecting mobile nodes, in specific area at specific time and managing its intrinsic characteristics, is challenging and require network designers to address the overall constraints in all network layers. Established routes are generally multi-hop due to limited transmission range of mobile nodes, so routing protocols must be able to route data packet through intermediate nodes until reaching targeted destination [35].

Secure data transfer and privacy protection are critical requirements for high risk operations like military and health operations. Communication Encryption is enabling confidentiality management but requiring high resources which are not available in WSN[36]. So, new techniques emerges to distribute and reduce computation requirements per node. Also detecting attackers is key for protecting the network and delivering the information's to the base station.

Here after we are going to highlight the main issues that impact network efficiency and scalability:

- **Energy-constrained**: Almost all WSN nodes rely on batteries or other exhaustible resources for powering up during the full device lifetime. Consequently, the most important system design criteria may be optimizing energy consumption. For most of the sensor nodes, communication functions should be optimized for lean power consumption. Conservation of power and power-aware routing must be taken into consideration[37] [38] [39].
- **Security and Reliability**: Beside common vulnerabilities of wireless connection, wireless sensor network has particular security problems due to nasty neighbor. The feature of distributed operation requires different authentication schemes and key management. Furthermore, wireless link characteristics introduce reliability problems, because of limited wireless transmission range; broadcast nature of wireless medium such as hidden terminal problem and data transmission errors. Mobile wireless networks are generally more prone to physical security threats than fixed networks. The increased possibility of eavesdropping, spoofing, and denial-of-service attacks should be carefully considered[40] [41].
- **Routing**: Owing to the dynamism of network topology, problems stack up in routing also. Since the nodes are continuously, table driven routing protocol can be used, therefore only reactive routing protocol can be used. Again multicast routing becomes challenging because nodes move freely and multi-cast tree is no more static. Also it is not necessary that source and destination nodes lie within each other's radio range, therefore multi-hop communication is needed, which is complex than single hop[37] [42].
- **Device discovery**: The identification of new nodes is a kind of tasks WSN should handle autonomously to ensure scalable network constructions and informing about the topology changes. Also, ensuring self-healing competences through detection of dead nodes and elimination of faulty nodes to address administration to maintain a dynamic update across the network to facilitate automatic selection of optimal routes.
- **Quality of Service (QoS):** Providing different quality of service levels in constantly changing environment is complex and challenging task. The inherent stochastic features of communications in WSN make it difficult to guarantee and deliver required SLR (Service level requirement) required by the application and use case of WSN.

Thus, adaptive QoS scheme must be implemented in addition to the traditional resource reservation techniques to support multimedia and real time services [43] [44].

- **Scarce Bandwidth**: The main limitation of wireless communication technique is variable link capacity that contributes to set up a path. Wireless links is going to still having significant lower end to end throughput than their hardwired counterparts [45].
- **Limited resources**: Not only power and bandwidth are scarce resources in WSN nodes, processing and memory are also very limited. The protocols and techniques available for fixed network infrastructures needs to be adjusted taking into consideration the limited computing capabilities.
- **Multicast:** Multicast is desirable to support multiparty wireless communications and guarantee efficient usage of scarce bandwidth resources. When nodes are moving, multicast tree is no longer static, multicast routing protocol must be able to cope with mobility including multicast dynamic membership (leave and join) [46].
- **Energy hole problem**: Nodes that are located on the boundaries of Sink or base station may suffer from excessive energy consumption meanwhile the geographic routing tends to delivers data packets along the hole boundaries by perimeter routing if it needs to bypass the hole[47]. This may enlarge the hole because of excessive energy consumption of boundaries nodes.

10. Conclusion

In a few years, advanced communication and networking technologies have occupied a part of all aspects of our life. This is noticeable in the way people interact, shop, do business, etc. This chapter presents a focus on wireless sensor networks. First, we presented different type of existing networks based on media type used to establish communication between a pair of source and destination node. Then, we detailed various infrastructures-less networks, which are divided into single-hop and multi-hop mode. Second, we give more details about the usage of wireless sensor networks. Finally, we discussed the main characteristics and various interesting application of WSNs. Moreover, we highlighted the main challenges that researches should address in order to open ups the door to a large-scale deployment. In the next chapter, we are going to study one of major issues in WSN, which are security threats and data privacy protection. This should be carefully designed to establish secure network that could fit to high requirements from a large amount of applications.

Chapter 2

State-of-the-art key disciplines in WSNs

1. Introduction

In recent decades, the development of mobile applications have been growing at a very high speed such as online banking, online gaming, E-learning, online trading, e-marketing and social media. The main driver behind fast penetration of mobile applications and virtual services is the standardization and the development of wireless technologies in particular wireless networks and the ability to use services from any device at any time and from any location. Thus wireless network, such as Wireless Sensor Network has become a success key of several humanitarians operations like rescue operation in battle-flied, disaster area …

As discussed in chapter 1, WSN has several intrinsic limitations that could be reviewed as follows: unreliable radio transmission medium, multi hop communication scheme, limited nodes' power and processing capabilities. Research community in wireless sensor network is mainly carrying active work in security fields, system capacity, power source lifetime optimization and compression. Because of the importance of security in critical wireless sensor network applications, various models and techniques to secure the information processing inside the network have been proposed to overcome network constraints and guarantee the level of data protection required by wireless sensor applications. There are some challenges that make the design of secure data collection in wireless sensor network a tough task:

- Restricted power of sensor nodes, computing capacity and scarce bandwidth require combination of efficient techniques for compression, encryption and routing schemes.
- The broadcast nature of wireless medium introduces active and passive attacks either inside or outside the network nodes;
- Deployment in harsh environment makes nodes vulnerable to attackers;
- Data packet may be lost due to variable and unpredictable capacity of wireless links;

The importance of considering many disciplines in providing secure data communication over wireless sensor network was the main motivation behind writing this chapter in which we aim to give overview on key techniques which contribute to secure communication taking into consideration WSN constraints. In the following, we propose to review data aggregation structures and their contribution to reduce energy consumption. In section 3, we will present the compressive sensing which enables distributing the compression efforts across all nodes in WSNs and discuss how this compression is possible in the context of WSNs and how energy consumption can be reduced. In section 4, we will study routing protocols in WSNs in order to understand the different strategies used in designing routing protocols and gain proper knowledge about WSN routing protocols. In section 5, we will present the security challenges and provide taxonomy of wireless sensor network attacks per communication layer. The attacks and security requirements are the pre-requisite state of the art in providing a secure design.

2. Data aggregation

WSN applications like environment monitoring and smart home systems are witnessing a quick deployment increase thanks to the cost reduction and technology reliability which are key factors for such success of WSN in these application areas. The network lifetime depends on the sensor nodes life, which is mainly related to energy consumption and availability of the sensor nodes. Data aggregation techniques are used to decrease communication overhead and thus decrease energy consumption which leads to an increase of sensor nodes life. Only aggregated data is being forwarded to the base station rather than sending all the raw data.

When multiple nodes are sensing same phenomenon, sensed data from the nodes is highly redundant in space or in time, and aggregation becomes a key technique to reduce the volume of useful data that need to be transferred to the base station. Figure 2.1 shows the working phases for data aggregation in wireless sensor network.

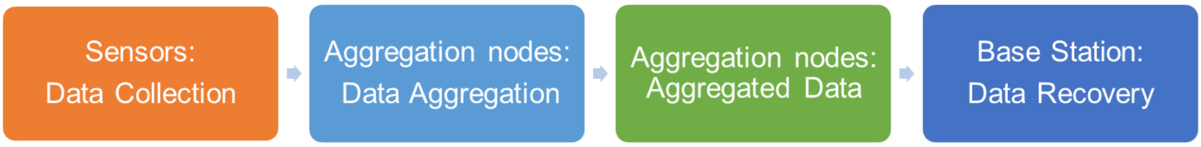


Figure 2.1 Data Aggregation phases in WSN

Data aggregation goal is to achieve the 3 following main objectives:

Saving Energy and Network capacity: by eliminating redundant or correlated data, aggregation will allow to reduce the overall raw data quantity that needs to be forwarded to the base station to a smaller quantity of useful data depending on the application. Reducing the volume of data means less communication over wireless and then less energy consumption.

Recovering Data: Depending on the aggregation function, the initial sensed data may be lost if simple functions are used like the average of reading values as received from leaf nodes. Data recovery requirements are related to the application and aggregation function should be designed to manage the recovery needs while saving network resources.

Protecting data privacy: Gateway nodes used for data aggregation are critical ones and are usually targeted in wireless sensor network attacks. Secure data aggregation and privacy protection is a key objective whenever data privacy is required by the application in scope.

Data aggregation techniques can be classified in 2 main categories:

Aggregation structure:

Defines the path of aggregation and network aggregators that will be gateway nodes transferring data from the leaf sensor nodes until the base station. The aggregation path s pre-defined when setting the network topology or updating it and is also a routing path in multi-hop WSNs.

Aggregation function:

The function used to aggregate data at sensor level is defining the aggregation function. Depending on the application, aggregation functions can be a simple ones like average, min or max. More sophisticated aggregation functions have been also evolved from research effort to improve ability to recover data with high accuracy.

Since the emergent of Wireless Sensor Networks, several aggregation structures and techniques have been developed in order to meet required functionalities related to specific application field. As a result, there is no aggregation technique that could fit to all wireless sensor networking contexts. Thus, establishing a classification of aggregation techniques according to the purpose or the goal for which the aggregation is designed will serve as reference model that help researchers make quick decision regarding the choice of the best aggregation technique to specific network context.

Let first discuss, how aggregation techniques makes possible data transfer and communication between two parties whether in wired or wireless networks and then present different criteria for classifying aggregation techniques in wireless sensor networks.

2.1 Data Aggregation Structures in Wireless Sensor Network

There are four main and commonly used data aggregation structures or methodologies in wireless sensor network named: Cluster based approach, Tree based approach, Centralized approach and In-Network aggregation as shown in figure 2.2.

2.1.1 Centralized Approach

In the centralized approach, each node is sending its readings and data to the same central node. And the aggregation is done at central node level before sending aggregated data to the base station. The central node should have enough energy and bandwidth to communicate with all nodes in the network. This structure is relevant for low density network and where distance is allowing direct communication with one central node.

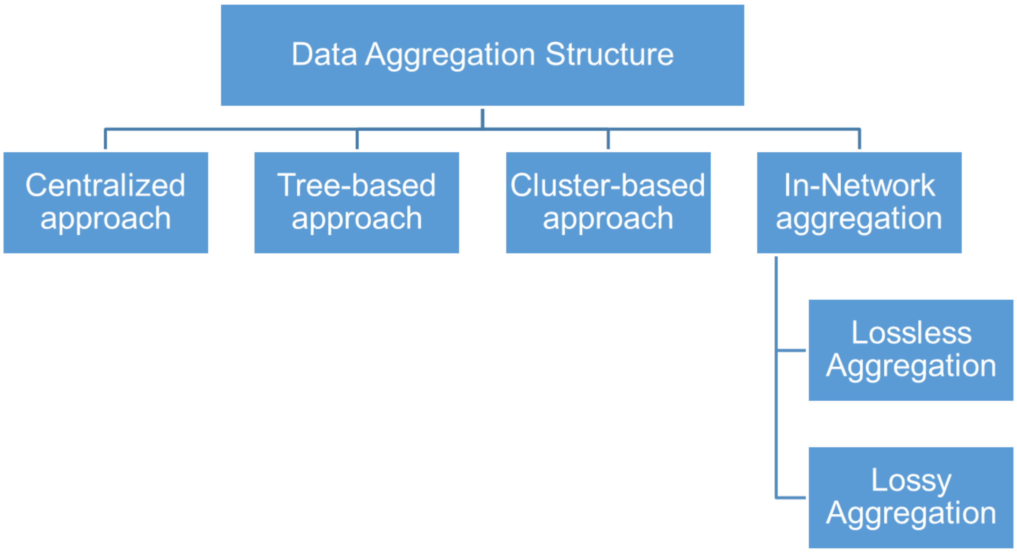


Figure 2.2 Data Aggregation structure

2.1.2 In-network Aggregation

In-network aggregation methodology is based on distributing the communication effort inside the network at nodes level which enable better energy management compared to centralized approach. This approach can be split into lossy and lossless methodologies. In some application, we can notice that there some special or temporal correlation between nodes readings and thus there is no need to send all data. For example in precision agriculture, the temperature will not change too much between 2 sensors in the same field. The average temperature and light intensity would be enough to monitor the field and decide if there is a need to adjust watering or the light of the monitored area. Depending on the application, function like count (), sum (), maximum (), minimum () or average () can be used. These function used for aggregation and only the aggregated sum for example is sen to the base station so it's a lossy aggregation methodology. Lossless data aggregation is not doing any compression nor using any aggregation function. All sensed data is sent to the base station in lossless data aggregation methodology.

2.1.3 Tree Based Approach

To form the tree, a minimum spanning tree is executed to find the route from leaf nodes or child nodes to the base station. Child nodes send the sensed data to the parent nodes and the aggregation is happening at parent nodes. A parent can be seen also as a child while forwarding data to next hop (parent node) until reaching the source node (base station).

In the following we will be interested into a

2.1.4 Cluster Based Approach

This approach is dividing the nodes into groups forming clusters. Each cluster has a Head of cluster which is responsible for data aggregation. Cluster Heads get data from cluster nodes and after aggregating data, they forward it to the base station.

2.2 Data Aggregation Techniques in Wireless Sensor Network

Data aggregation techniques are various and each technique intend to focus on optimizing one or multiple factors. Data Aggregation Techniques can be focusing on energy, network lifetime, network density, Quality of Service or latency and the name of the technique is associated to the same like Scheduling based technique which is using scheduled communication to optimize the aggregation. Cluster based techniques are particularly interesting in the context of Wireless Sensor Network. Data Aggregation processing can use different techniques depending on the two types of the network namely Hierarchical or Network Flow based as shown in Figure 2.3

LEACH protocol is an example of cluster based network which fall under the hierarchical structures. In cluster based data is transmitted from cluster nodes to cluster heads within the same cluster and then after aggregation, data continue to flow from one cluster to another cluster through cluster heads until the base station.

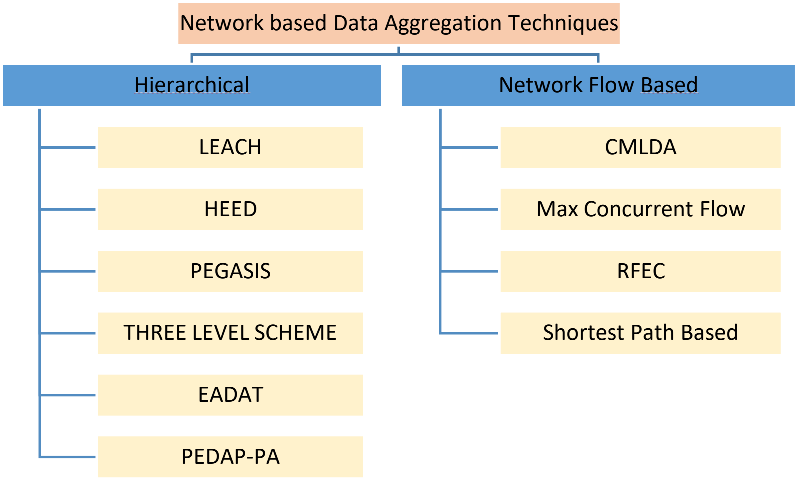


Figure 2.3 Network based data aggregation techniques

Second hierarchical network is chain based in which, data transmission is done from node to closer neighbor node after the structuring of nodes into linear chain using greedy algorithm for data aggregation (Example: PEGASIS). The third and last hierarchical network is tree based, in which all the nodes are structured in the form of tree and transfer the data from leaf node to root node through intermediate node for data aggregation (Example: EADAT).

Technique	Objectives	Organization type	Features
LEACH	Network lifetime is number of alive nodes	Cluster	Randomized rotation of cluster head (CH) Non-uniform energy consumption in various sensors
HEED	Network lifetime is number of rounds till the death of first node	Cluster	Many energy levels in sensors Performs better than LEACH CHs are well distributed
PEGASIS	Network lifetime is average of energy consumed by a node	Chain	Requirement of overall knowledge of network Less energy consumption as compared to LEACH
CHAIN BASED	Network lifetime is multiplication of power consumption and delivery delay	Chain	Three level scheme better than PEGASIS and Binary chain based scheme better than LEACH in terms of performance
EADAT	Network lifetime is number of live sensor nodes till complete transmission of data	Tree	Sink initiated broadcasting technique For broadcasting help messages select the threshold power
PEDAPPA	Network lifetime is till the expire of last node	Tree	Minimum spanning tree based technique Performs better than LEACH, PEGASIS

Table 2.1 Hierarchical data aggregation techniques. Source[26]

The network flow based data aggregation is designed to optimize the network flow while aggregating data. The technique is designed to increase network lifetime and reduce energy consumption along with optimizing flow constraints on information.

[26] Sukhchandan Randhawa, Sushma Jain, Data Aggregation in Wireless Sensor Networks: Previous Research, Current Status and Future Directions, Springer, july 2017

In the following section, we will detail the LEACH protocol as a cluster based technique which is adapted to the wireless sensor network.

2.2.1 LEACH Protocol

LEACH (Low Energy Adaptive Clustering Hierarchy) is one of the main proactive protocols in wireless sensor network, it is a hierarchical self-organizing routing protocol based on adaptive clustering used in WSN to minimize the energy consumption of network nodes in order to increase the lifetime of the network [48] [49]. LEACH was proposed by Wendi B. Heinzelman of MIT [50]. LEACH assumes Base Station is not moving and located far from sensor nodes. In addition, all the sensor nodes are homogeneous and have limited energy and memory. Sensors can communicate with each other and they can communicate directly with the Base Station (BS). The main idea of the LEACH protocol is to organize the nodes in clusters to distribute the energy between all the nodes of the network. Thus, in each cluster, there is a master node called Cluster Head (CH) which gathers the data received from the sensors of its cluster and transmits them to the base station which allows to minimize consumption and reduce the amount of information sent to the base station. The LEACH protocol is executed in several cycles (or rounds). Each round is composed of two Phases: the configuration or initialization phase (setup phase) and the transmission phase (steady-state phase). The round is made up of several frames, each frame is made up of slots [51].

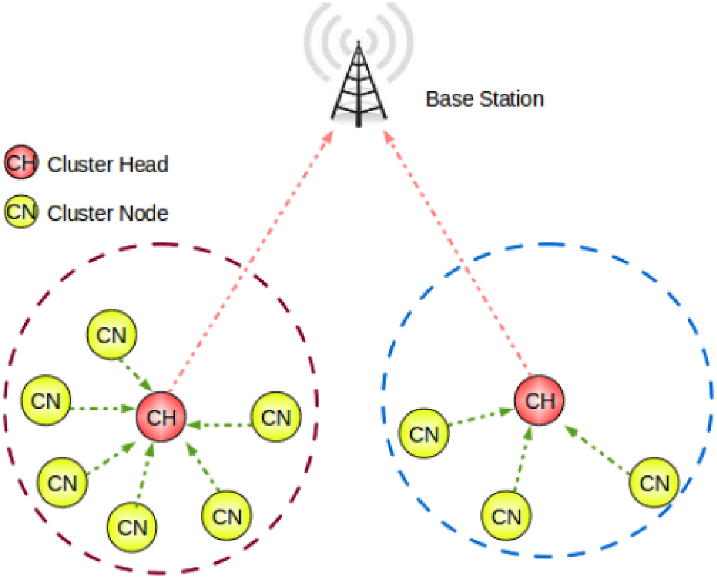


Figure 2.4 Cluster based Network and routing topology

Steady-state phase or the transmission phase is longer and consumes more energy than the configuration phase (set-up phase).

Figure 2.5 shows the two execution phases of a round for the LEACH protocol.

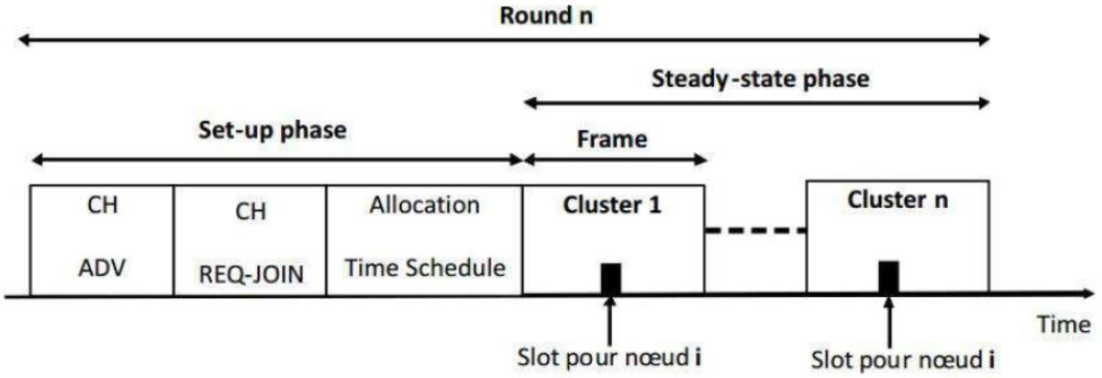


Figure 2.5 Two phases execution for LEACH protocol.

2.2.1.1 LEACH Protocol Setup Phase:

At the start of the round, each node decides independently of the other nodes to become or not a cluster master for the current round.

Each sensor node generates a Random number such that 0 <Random <1 and compare it to a predefined threshold T (n).

if Random <T (n), the sensor node becomes Cluster Head during this round, otherwise it is a member of the Cluster [50].

The threshold T (n) is given below:

$$\mathrm{T(n)} = \begin{cases} \mathrm{T(n)} = \frac{P}{1-P\left(r\ mod\left(\frac{1}{P}\right)\right)}\ , & \text{if n } \in \mathrm{G} \\ 0 & ,\ \ Otherwise \end{cases} \tag{2.1}$$

Where,

P: the probability that the node is selected as a cluster-head. In LEACH, P is also the percentage of the nodes acting as CH. The choice of P depends on the network density to have the adequate number of CH.

n: the number of given nodes.

r: the number of current round

G: the set of nodes that have not been CH during the last 1/ P rounds

The nodes which are CHs during round r must not be selected during next 1/ P rounds. After 1/ P -1 rounds, the threshold value becomes 1 for any sensor node which has not yet been CH, and after 1/ P rounds, all nodes are eligible again to become CHs. Once the CHs have been allocated to all the clusters, each CH will disseminate an advertisement message (ADV CH) to the rest of the nodes that it is the new cluster head using the Carrier Sense

Multiple Access - Media Access Control (CSMA-MAC) protocol [52]. Each non-CH node, after receiving the advertisement message, choose the appropriate (closest) cluster or CH to join. This selection is based on the signal strength of received RSSI (Received Signal Strength Indication) messages. Through subsequently, each non-CH node informs the CH of the chosen cluster by sending a request to join containing their identity using CSMA (Carrier Sense Multiple Access) to join the cluster (JOIN-REQ) [53].

During the configuration phase, all CHs keep their receivers on. After the establishment of the clusters, each CH creates a temporal sequence (TDMA table), according to the number of nodes of its cluster. CH communicate with cluster members (CM) and distribute to each member node of the cluster the appropriate time slot when it can transmit.

TDMA scheduling prevents collision between data messages and save energy between the non-cluster nodes. Thus, all member nodes know their TDMA locations, the principle of this TDMA technique consists in allocating the total bandwidth at the node during a given slot time. Each CH plays the role of the coordinator to control the transmissions of the members thanks to the sending prerequisite of the TDMA table to members of their cluster. So this allows each member of the cluster knows their time slot that it will occupy, and to pass to the state slept during inactive slots [54] [55].

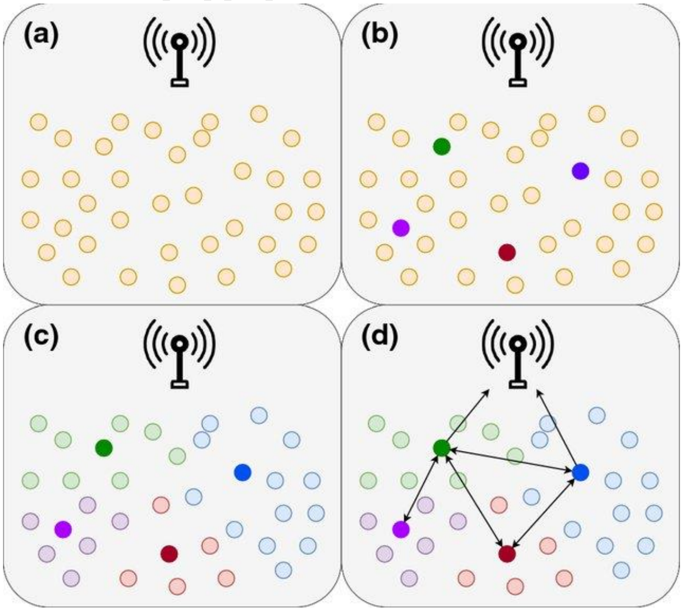


Figure 2.6 LEACH Protocol, a-b-c indicate the setup phase and d shows the steady phase. Source[27]

- [27] EDMARA2: a hierarchical routing protocol for EH-WSNs, Milad Khademi Saeed Sharifian, Springer, August 2020

Figure 2.6 illustrate the phase of Leach protocol, in step a, nodes are choosing their random numbers and start communicating with their neighbors. In step b and c, nodes elected as CHs and nodes are joining one cluster each. Step d is related to steady-state phase where communication start for N rounds where N is the number of nodes in the network.

2.2.1.2 LEACH Protocol Steady-State Phase:

In the transmission phase, each sensor node member of the cluster detects the environment and transmit the sensed data to their CH according to the TDMA program. Member nodes go into sleep mode to save power. When the CH receives all the data sent by the members of its cluster, it aggregates them then send to BS.

After a given time interval, a randomized rotation of the role of CH is conducted so that the uniform consumption of energy in the sensor network is obtained. LEACH uses the CDMA multiplexing technique for the CHs to send the data to the base station, where inter-cluster communication will take place. Communication can be direct (a single jump) or indirect (multi-hops). A new round will have to take place after the end of this phase, This process is repeated until all the nodes of the network are elected CH, only once, throughout the previous rounds. In this case, the round is reset to 0.

HEED (Hybrid Energy Efficient Distributed clustering approach [56] is based on LEACH and aims to improve energy consumption. In HEED, the head cluster selection is not random anymore where every node should become cluster head once during the N rounds where N is the number of nodes. The HC probability of selection is linked to the remaining energy available for each node. Each node calculate its probability of becoming CH, P_{CH} as follows:

$$P_{CH} = P.\frac{E_{residual}}{E_{max}} \qquad (2.2)$$

Where,

P : The initial percentage of CHs,

$E_{residual}$: The residual energy of node,

E_{max}: The initial energy (basically it's the maximum energy of the node).

HEED is very similar to LEACH and uniforms the CH selection based on the energy consumption. The issue with HEED is that nodes with lower energy E_{max} will consume their energy faster. So, the designers of the Wireless Sensor Network should take into consideration the protocol constraints which is more suitable for networks where E_{max} is same for all nodes and there is no expansion of the network during time which can lead to high difference in E_{max} between sensor nodes.

2.3 Data Aggregation Taxonomy

The data aggregation taxonomy are key indicator used to measure the performance of data aggregation techniques and evaluate their suitability to the applications. The mainly used taxonomy indicators are Network lifetime, energy efficiency, data accuracy, latency and data aggregation rate.

Energy Efficiency: indicate the ratio between nodes functionality and its consumed energy to achieve its role. Data aggregation technique is considered energy efficient when it delivers the highest functionality with lowest energy consumption in WSNs. Energy efficiency for the WSN is the sum of energy efficiency of each node i where n is the total number of sensor nodes in the network.

$$\text{Energy Efficiency}=\sum_{i=1}^{n}\left(\frac{\text{Amount of data } successfully \text{ transferred by sensor node i}}{\text{Total energy consumed to transfer those data}}\right)(2.3)$$

Network Lifetime: This is the lifetime of the first node to quit the network because it runs out of energy. So, Network lifetime can be defined as the time or number of rounds until the first sensor node or group of sensor nodes in the network exhausted of its energy. Network Lifetime is not fully an indicator of global network lifetime as this depends on the size of the network and its density. In high density network, several nodes can be dead but the network still continue to operate. In low density WSN, Network Lifetime NL^n can be reached if one or few sensors are not active anymore. In all case, this indicators give good estimation about the overall network lifetime when combined with the density and application requirements.

$$NL^n = \min_{v \in V} NL_v \qquad (2.4)$$

Where NL_v is the Network Lifetime of node v and V is the set of all nodes excluding the Base station.

Data Accuracy: is related to the ratio of data successfully transferred in the network by all nodes to total data transferred by every node. Data accuracy shows the reliability of the network and ability to deliver data without failures.

$$\text{Data Accuracy} = \sum_{i=1}^{n} \left(\frac{\text{Amount of data successfully transferred by sensor node i}}{\text{Total amount of data sent by sensor node i}}\right) \qquad (2.5)$$

Latency: is the time required for the data to be transferred from sensor nodes to the next node. The total of the transfer duration for all transferred data is defining the overall network latency. Equation 2.6 defines latency as a time difference between sending data by a sensor A and receiving the same data by a sensor B.

$$\text{Latency} = \sum_{i=1}^{n} (\text{Time of receiving data} - \text{ Time of sending data}) \qquad (2.6)$$

Data Aggregation Rate: This indicator shows how much compression or data reduction ratio the aggregation function and technique is able to achieve. High data aggregation rate is key to reduce energy consumption and save energy in Wireless Sensor Network.
Data aggregation rate at aggregator node level or Cluster is the ratio of quantity of data aggregated successfully to total quantity of data sensed by all sensor nodes within same cluster. The total data aggregation ratio is the sum of aggregation ratios for all clusters.

$$\text{Data Aggregation Rate} = \sum_{Clusters} \frac{\text{Amount of data aggregated successfully}}{\text{Total amount of data sensed}} \times 100 \qquad (2.7)$$

3. Compressive Sensing

Compressed sensing is a new approach of compressing signals. The classical way of compression proceed by getting full information’s and then keep only the useful one. So there is a waste of time and resources to capture all information’s and then get rid of information’s that are not required to recover the signal during the decompression phase.

Compressed sensing suggest sensing only the useful information from a compressible signal that means acquiring a small number of non-adaptive linear measurements of the signal. Just the process of sampling can compress the signal. The decompression phase uses different algorithms taking advantage from signal sparsity.

The work results of Kotelnikov, Nyquist, Shannon, and Whittaker [57] [58] [59] demonstrate that signals, images, videos and other data can be recovered from a set of uniformly spaced samples taken at the so-called Nyquist rate of twice the highest frequency present in the signal of interest.

Unfortunately, in many emerging applications such as medical imaging, video, remote surveillance, resulting Nyquist rate is so high that the data samples is too high.

The fundamental idea behind CS is rather than sampling at a high rate and then compressing the sampled data, we would like to find ways to directly sense the data in a compressed form – ie at a lower sampling rate. In their works[60], Emmanuel Candès, Justin Romberg, and Terence Tao and of David Donoho, demonstrate that a finite dimensional signal having a sparse or compressible representation can be recovered from a small set of linear non-adaptive measurements [61].

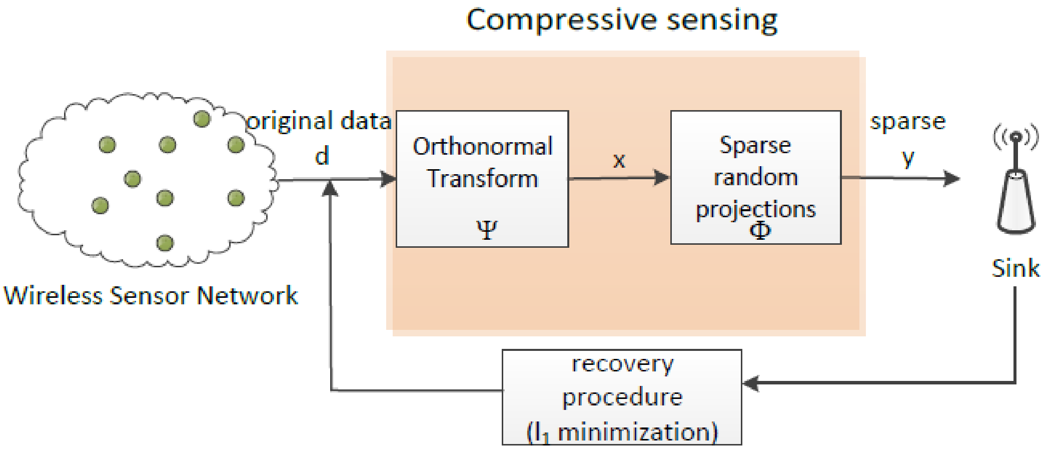


Figure 2.7 The procedure of compressive sensing in Wireless Sensor Network

Figure 2.7 shows the procedure of compressive sensing in wireless sensor network.
A sensor node reading are designed as the original data d. Compressive sensing is processed at node level via multiplying d with orthonormal matrix Ψ and random sparse matrix Φ . The compressed data y is then forwarded across the network until the sink node.
Recovery of Raw data d at sink node is costly in term of resource and energy. This is not an issue as sink node is generally having enough resources. The recovery procedure is achieved via a minimization as described in the following section.

3.1 Compressed Signal Recovering

The fundamental idea behind Compressed Sensing is rather than sampling at a high rate and then compressing the sampled data, find ways to directly sense the data in a compressed form at a lower sampling rate.
Recovering a signal $x \in \mathbb{R}^N$ from $y \in \mathbb{R}^M$ measurements such as M<<N. Φ is a measurement matrix

$$y = \Phi x \tag{2.8}$$

Where M equations and N unknown. There are much more unknown than equations. So there are many solutions $x \in \mathbb{R}^N$ with $y = \Phi x$ without additional hypothesis about x we cannot recover x from y and Φ with M<N

By a compressible representation, we mean that the signal is well approximated by a signal with only few nonzero coefficients. To consider this mathematically, let x be a signal that is compressible in the basis Ψ :

$$x = \Psi\alpha$$

Where α are the coefficients of x in the basis Ψ . The error between the true signal and its K term approximation is denoted the K-term

$$\sigma_k(x) = \arg\min_{\alpha \in \sum_k} \|x - \Psi\alpha\|_2 \tag{2.9}$$

Approximation error $\sigma_k(x)$, defined as for compressible signals, we can establish a bound with power law decay as follows:

$$\boldsymbol{\sigma_k(x) \leq C_2 K^{1/2-s}} \tag{2.10}$$

In order to recover a signal x from measurements y, we should answer two main questions:
Question1: How should we design the sensing matrix Φ where $y = \Phi x$ to ensure that it preserves the information in the signal$x$?
Quastion2: How can we recover the original signal x from measurements y?

3.2 Sensing matrix design

Null space condition: The sensing matrix Φ uniquely represents all the sparse signal $x \in \sum_k$ if and only if the null space of Φ denoted $\mathcal{N}(\Phi)$ contains no vectors in $\sum_{2k}$. $\mathcal{N}(\Phi)$= $\{z : \Phi z = 0\}$
In order to recover all sparse signals x from the measurements $y = \Phi x$, then for any pair of distinct vectors x and $x' \in \sum_k = \{x \in \mathbb{R}^N : \|x\|_0 \leq k\}$, we must have $\Phi x \neq \Phi x'$ which means if $\Phi(x - x')$=0 with$(x - x') \in \sum_{2k}$, then it's clear that Φ uniquely represents all $x \in \sum_k$ if and only if $\mathcal{N}(\Phi)$ contains no vectors in $\sum_{2k}$.

Which means there exist no vector Z which is 2K sparse (contain a maximum of 2K non-zero coefficients) such that $\Phi z = 0$

3.3 Sparse Signal Recovery via l1 Minimization

Given measurements y such that $y = \phi x$ and the knowledge that our original signal x is sparse or compressible, the first option to consider is trying to recover x by solving the optimization problem of seeking the sparsest signal: $\hat{x} = \underset{z}{\operatorname{argmin}} \|z\|0$ such that $z \in \mathcal{B}(y) = \{z\colon \Phi z = y\}$; if the signal is noise free

Dealing with $\| . \|0$ which is a non-convex function, is very difficult to solve as an optimization problem. In fact, for a general matrix ϕ, even finding a solution that approximates the true minimum is NP-hard.

A tractable solution of this problem can be achieved by replacing $\| . \|_0$ with its convex approximation $\| . \|_1$ which is computationally feasible.

$$\hat{x} = \underset{z}{argmin} \|z\|_1 \ \ Subject\ to\ z \in \mathcal{B}(y) \tag{2.11}$$

3.4 Compressive Sensing based Techniques

The work in [62] proposes a Compressive Sensing data gathering algorithm based on Packet Loss Matching (CS-PLM) to maximize the accuracy of data reconstruction. In a tree-based routing structure, the unreliable communication links impose huge packet loss in the network, resulting in reduced data reconstruction accuracy of compressive sensing based data collection. To avoid such an issue, the CS-PLM constructs a sparse matrix that is estimated by matching the packet loss in the network. Consequently, the CS-PLM checks the sparse matrix attains a probability value close to 1 by satisfying the Restricted Isometry Property (RIP). Moreover, the CS-PLM ensures reliability in the data forwarding by enabling a multipath routing technique among compressive sensing nodes. The work in [63] proposes a cloud-assisted compressive sensing model to gather data securely. In such a model, three various parties, such as sensor, cloud, and user, involve in the algorithm process and offer several advantages. To attain better tradeoff between data forwarding reliability and energy dissipation, a novel method, named as a spatial-temporal compressive data gathering algorithm (ST-CDGA), has been proposed in [64]. The ST-CDGA consolidates a Kronecker compressed sensing (KCS) with a cluster structure and also employs the correlations of Spatio-temporal measurements concurrently. By enabling the cluster head nodes to construct a matrix based on sparse sub-measurements retrieved from data from nodes that are not able to forward their measurements directly, ST-CDGA improves the data reconstruction accuracy. Further, the base station in ST-CDGA exploits the spatial correlation of cluster topology and involves forming a block diagonal matrix (BDM).

The work in [65] presents a framework named adaptive compressive sensing to monitor the environmental conditions periodically. The adaptive framework also constructs two modules that are reconstruction error and sparsity identification. The reconstruction error module is used to verify that the current sampling rate is enough for reconstructing the signal, and the sparsity identification is utilized to evaluate the sparseness of signals over a specific time interval. The work in [66] applies an energy-efficient data collection method, named as Cluster-Based Compressive Sensing Data Collection (CCS) over WSNs. The CCS model integrates the clustering topology with block-wise compressive sensing, which utilizes the sparse sensor readings based on spatial correlations for compressive sensing. The CCS applies the compressive sensing at the cluster members only, and the data is recovered by CH nodes before CH to base station data forwarding. The CH recovers the data directly or indirectly, named as DCSS and ICSS, respectively. Moreover, the CCS minimizes the data transmission cost over WSN significantly.

4. Routing in Wireless Sensor Network

The main building bloc in networking is routing. Therefore, designing routing protocols have attracted the interest of researchers. Since, several routing protocols have been proposed in order to meet required functionalities related to a specific application field. As a result, there is no routing protocol that fit to all ad hoc networking conditions [67][68]. Routing protocols can be classified using several approaches, depending of the purpose or the goal for which the protocol is designed. See illustration in figure 2.8. There are different criteria for classifying routing protocols in ad hoc networks:

- Communication Model;
- Network structure;
- Scheduling model;
- State Information;
- Route establishment;
- Type of Cast;
- Type of path.

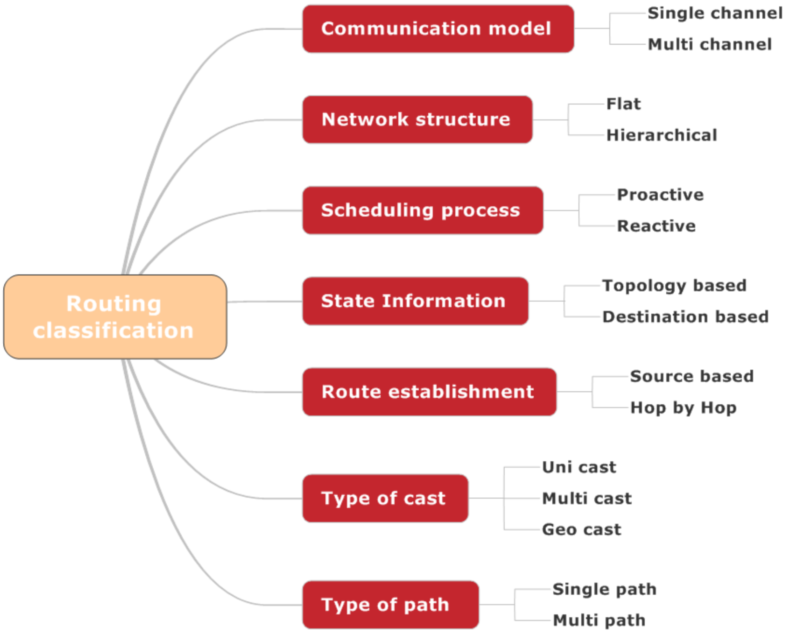


Figure 2.8 Routing classification

4.1 Communication model

Routing protocols can be designed to work in different wireless communication schemes. Wireless communication models can be single channel or multi-channel [69]. First, single channel schemes were designed for Medium Access Control (MAC) to address physical layer deficiencies and provide reliable information to upper layers, these schemes suffer from hidden and exposed terminal problems, as illustrated in figures 2.10 and 2.11, fairness and power consumption issues related to radio communication in wireless sensor networks. Most of designed protocols have partially solved the intrinsic problems of wireless communication. On the other hand, multichannel schemes have shown better capabilities to handle hidden and exposed terminal problems due to the usage of more than one channel in their network. Multichannel protocols are generally used in Time Division Multiple Access (TDMA) or Carrier Sense Multiple Access (CSMA) based networks. In contrast, single channel are Carrier Sense Multiple Access with Collision Avoidance (CSMA-CA) scheme based.

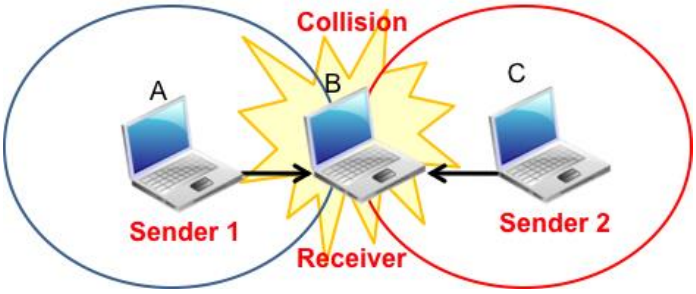


Figure 2.9 Hidden node problem

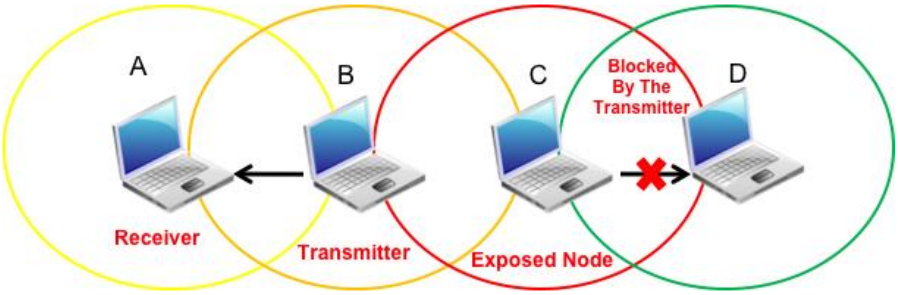


Figure 2.10 Exposed node problem

4.2 Network structure

Depending on how nodes participate in routing task, routing protocols can be considered flat or hierarchical [70]. In flat protocols, all nodes have the same responsibilities in the entire routing process [71]. Hence, control messages of routing are globally managed in uniform manner, which cause scalability problem in large networks. For example, Destination sequenced Distance Victor (DSDV) and Ad Hoc On demand Distance Victor (AODV). In hierarchical protocols, the main concern is reducing routing control messages for scalability purposes [72]. For example, nodes in Zone-based Hierarchical Link State (ZHLS) are dynamically organized into clusters [73]. In hierarchical scheme, only cluster head nodes know topology information. Other nodes just send data to the cluster head, which in turn will execute the entire routing process such as finding optimal path to the destination.

4.3 Scheduling model

In the literature, most routing protocol classification in wireless sensor network are based on their route establishment and maintenance strategies [4][74][75][76]. Routing protocol is considered proactive or table-driven if nodes maintain route information all the time to all destinations. Thus, when a source node has data to send it looks up its routing information database and start immediately sending data to the destination, which guarantee lower transmission delay. The drawback of proactive routing is periodic routing table updates, which cause routing overhead problem. In reactive or on demand routing protocols, like AODV, route information is acquired and maintained only when source node has data traffic to send. Generally, reactive protocols have two main operations: the first operation is route discovery, which is initiated when a source needs route to destination. The second operation is route maintenance, which consists of repairing failed links through active route due to topology changes.

4.4 State Information

Routing protocols could be described in terms of information state acquired at each node. Two main categories are distinguished: topology based protocols and destination based protocols that are broadly used in traditional wired routing protocols. In topology-based protocols [77], nodes maintain global view of network topology [78]. This approach is known as link State, where link sate packets (LSP) are exchanged between all network's nodes. Every node constructs and maintains global network topology from received LSPs, and computes the best routes to all other nodes using Dijkstra's algorithm. In destination based protocols [79], nodes do not maintain large scale topology information. The main destination based protocols are Distance victor where every node periodically exchanges distance vector with its neighbors. When a node receives distance vector information, it computes new routes and updates its distance vector database. The complete path then established, in a distributed scheme, by combining the next hop of nodes on the path from source to destination node. Distance vector routing protocols have less computational complexity and overhead messages.

4.5 Route establishment

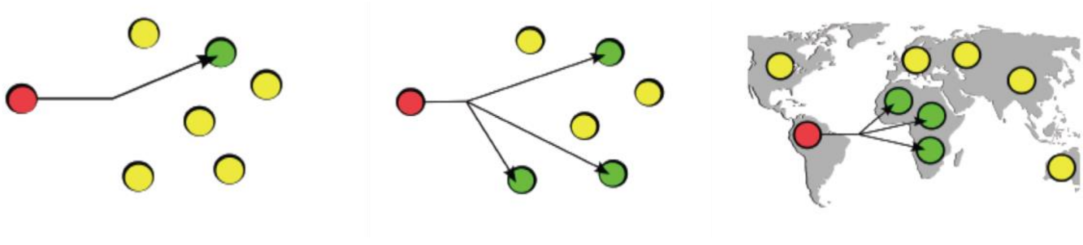
Routing protocols can be distinguished according to the way data packets are forwarded from source to destination node. There are two approaches [2][4][80]: First, source routing protocols, such as Dynamic Source Routing (DSR), which place the entire route information into packet header, then intermediate nodes just forward data packet according to route information stored in the header. In this approach intermediate nodes do not need to compute and maintain updated routing information, as a result much less time is needed for traffic delivery and much less control traffic is generated. However, Source routing do not scale very well in large network and dynamic topology, especially when the route is too long, data packet header become large and consume too much of scarce bandwidth. Second approach is hop by hop, which use next hop information stored at each node involved into an active path, like

OLSR. Thus, when a node receive data packet, it lookup the routing table and forward the packet to the next hop. The advantage of this strategy is that routes are adaptable to dynamically changing environments. The drawback of hop-by-hop routing is that each intermediate node has to maintain routing information for each active route and each node may require being aware of their surrounding neighbors through the use of beaconing messages.

4.6 Type of cast

Another way to classify routing protocols could be based on type of cast. For example: OLSR, AODV and DSDV. Unicast Routing Protocol is the most developed for MANET applications. In unicast routing one separate copy is sent to each receiver from the source node. Thus, data packet is replicated at source node and then delivered to all destination nodes; see figure 2.5.a. Unicast process consumes more much bandwidth due to redundant data packets. Multicast routing protocol, like On demand Multicast Routing Protocol (OMRP), has become very important in multimedia communications. To send simultaneously the same data packet to multiple receivers, the simplest way is broadcasting. However, broadcast technique consumes considerable bandwidth and power. Consequently, broadcasting should be avoided as much as possible in wireless sensor networks due to scarce bandwidth and limited nodes' energy. In multicast process, network components replicate data packet, which lead to optimal use of scarce bandwidth [81]. See figure 2.5.b.

Another type of routing protocols is geo-cast routing protocols illustrated in figure 2.5.c. This routing scheme has been adopted in VANET routing; it consists of sending data packet to a set of nodes inside a specific geographical area[82].



(a) Unicast

(b) Multicast

(C) Geo-cast

Figure 2.11 Type of cast

4.7 Type of Path

Some routing protocols are able to find multiple paths to a destination, like Multipath Ad hoc on demand Distance Victor protocol (AODVM), which make routing efficient in case frequent links break occur due mobility. In contrast, others routing protocols are simple and find only one path to the destination. Single path routing protocols should re-compute new route each time a link failure is detected, which become more complicated in highly dynamic environment [83].

5. Security in Wireless Sensor Network

Security in general and data privacy protection is an essential non-functional and sometimes functional requirement of most systems. Security is linked to the prevention of different types of unauthorized access to system functions or detection and prevention of attempts to block authorized access such as Denial of Service attack (DoS attack). Security must be taken into account from the first stages of network design in order to secure transmitted data and protect the network resources from unauthorized access [84]. Security protocols, methods and models used in wired networks are unsuitable for wireless sensor networks because of their limited energy resources and the nature of wireless communication.

The rapid development of wireless communication technology and small sensors allowed WSNs to spread rapidly and also to become a privileged target for attackers.

5.1 Security requirements in WSNs

WSN are vulnerable to attacks exploring the wireless transmission medium and its broadcast nature. The energy and resource limitations are also additional weaknesses on the sensor side. Sensor devices are usually deployed in uncontrolled environment, thus they are exposed to physical attacks. Security requirements in WSNs are similar to those of ad hoc networks due to similarities between MANET and WSN.

A sensor network is a special type of network, even if it’s sharing common properties with typical computer networks, which faces unique conditions because of its Specific characteristics. Thus, WSNs also have following general security requirements [85], namely:

- **Availability**: The service offered by whole WSN, by any part of it, or by a single sensor node must be available whenever required.

- **Integrity**: This is ensuring that the message or the entity under consideration is not altered. An intruding (adversary) node can modify the transferred data. For example, a malicious node can add some fragments or manipulate the data in a package. This new packet can then be sent to the original receiver. Data loss or damage can even occur without the presence of a malicious node due to harsh communication environment. Thus, the integrity of data ensures that no data received has been changed in transit.
- **Confidentiality**: it's about providing privacy of the wireless communication channels to prevent eavesdropping. Confidentiality is a very important security mechanism in wireless communication. In WSNs, it refers to the limitation of access to information to only authorized nodes and users who can access the data exchanged in the network and prevent unauthorized attackers. The main technique to ensure confidentiality is the usage of encryption.
- **Authentication**: it's related to authenticating other nodes, cluster heads, and base stations before granting a limited resource, or revealing information. It is the mechanism of verifying the identity of a node that wants to communicate with other nodes. Sometimes an attacker can forge and inject fake packets in the network, in this case the sensor node must be able to check whether the source node has valid identity.
- **Data Freshness**: Data freshness means received data is recent and not outdated. Old data could mean a man in the middle attack is ongoing and delays are related to data processing by the adversary node. Freshness can't be replaced by confidentiality and data integrity as for real time application freshness is a must.
- **Non-reputation:** preventing malicious nodes to hide their activities. Data transmission from one node to another should not be denied.

5.2 Attack classification in WSNs

According to specific criteria's, such as the power of the attacker, the belonging or not of the attacker to the network, attacks against sensor networks can be classified according to the following categories:

Passive Attacks:

Passive attackers are global eavesdroppers that can be placed in the network and observe part or even the whole communication of the network. Thus, they can access to all traffic and conduct statistical analysis based on the eavesdropped messages. By analyzing the traffic across the WSN, it is possible for a passive attacker to trace flow packets in the network. Passive attacks are only interested in collecting sensitive information without any modification or influence on communication. These collected information's as the detection of important nodes in the network (Cluster-Head) can then help the attacker to carry out malicious attacks.

Active attacks: attackers can forge, reply and modify messages. Especially in WSNs, attackers, such as bogus and replayed routing information attack, sinkhole and wormhole attack, selective forwarding attack, HELLO flood attack, and Sybil attack, can trigger various types of active attacks. The attacks activities have as object, the disturbance of the function of the network and the deterioration of its performance. Attacker tries to exploit security vulnerabilities on the network to launch various attacks in order to modify the data.

Node compromising attacks: this occurs when the attackers physically compromise sensor nodes, thus they can take control of it and the secret information stored on the sensor. Intrusion detection system can detect the compromised node if they have suspicious behavior. Detecting compromised sensors that still act normally is infeasible in all existing detection mechanisms in WSN.

Internal vs. External attacks:

An external attack occurs from outside the sensor network. That is to say they occur by nodes which are not deployed inside the network and that are not allowed to participate in the network. While the internal attacks occur by malicious internal nodes.

In all previous attacks scenario, we assume that the attacker have no computational nor energy constraint. Therefore, the attacker can proceed with all type of attacks.

5.3 Attacks targeting WSNs

The protocol stack [86] used by the sensor nodes comprises five layers which are: Application layer, transport layer, network layer, data link layer and the physical layer which have the same functions as those of the OSI model. Attacks are designed to take advantage from weaknesses or vulnerabilities related to a specific protocol layer and hence their classification. In addition to layers previously cited, the stack groups together three planes, namely the management plan of energy, the mobility management plan and the task management plan. Following the functionality of the sensors, various applications can be used and built on the application layer. The transport layer helps to manage the data flow if the network number of sensors requires it. It allows to divide the data coming from the application layer in segments to separate them, so it reorders and gathers the segments coming from the network layer before sending them to the application layer. The network layer takes care to route the data provided by the transport layer. MAC (Media Access Control) protocol, of the link layer, manages access to the physical medium and controls communication errors. The physical layer aims to meets the needs of a simple and robust modulation, and ensures the transmission and reception of data at bit level. In addition, energy, mobility and task management plans monitor and manage energy consumption, movement, and distribution of tasks between nodes sensors. These plans help the sensor nodes coordinate detection tasks and reduce overall energy consumption.

Energy is a scare resource in wireless sensor network and thus many attacks are targeting to exhaust battery power for sensors. About 80% of the energy consumption is happening during transmission and reception while wireless link is active. So attacks are focusing on keeping victims in this modes instead of sleep mode which is a saving energy mode.

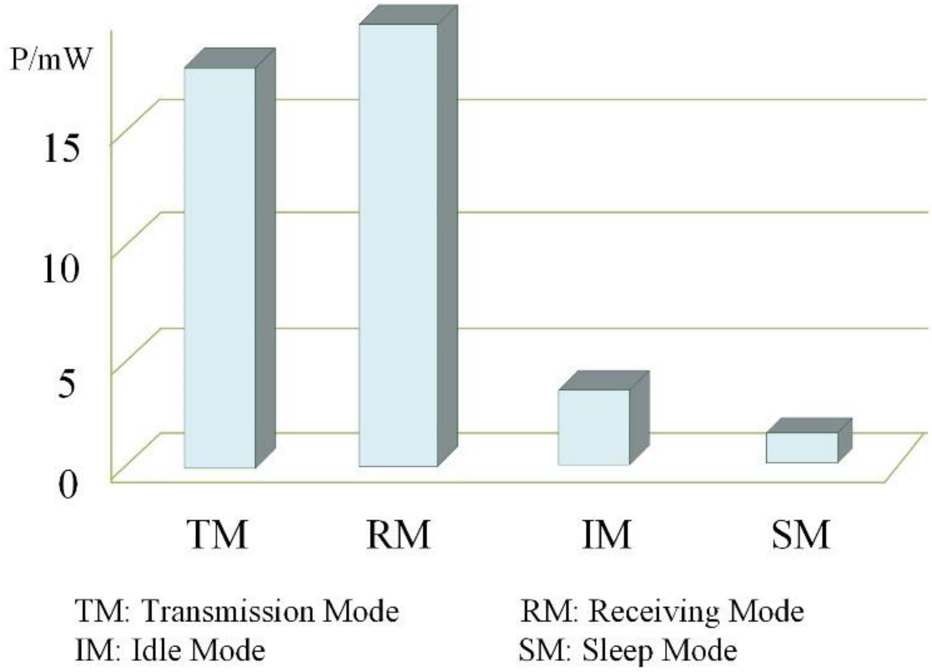


Figure 2.12 Energy consumption of different modes. Source[28]

[28] Data Collection for Security Measurement in Wireless Sensor Networks: A Survey, IEEE Internet of Things Journal (Volume: 6, Issue: 2, April 2019)

5.3.1 Attacks in Physical Layer

5.3.1.1 Eavesdropping Attack

In this attack, the attacker intercepts radio signals without destroying their integrity. Eavesdropping attack is a passive one that is considered as the first step for information's collection before launching other active attacks[87]. The attacker monitors message transmission and intercepts it. Non encrypted messages, can easily be read by the attacker. Like other passive attacks, it can't be detected especially if the attacker keeps listening without sending anything.

5.3.1.2 Compromised Node Attack

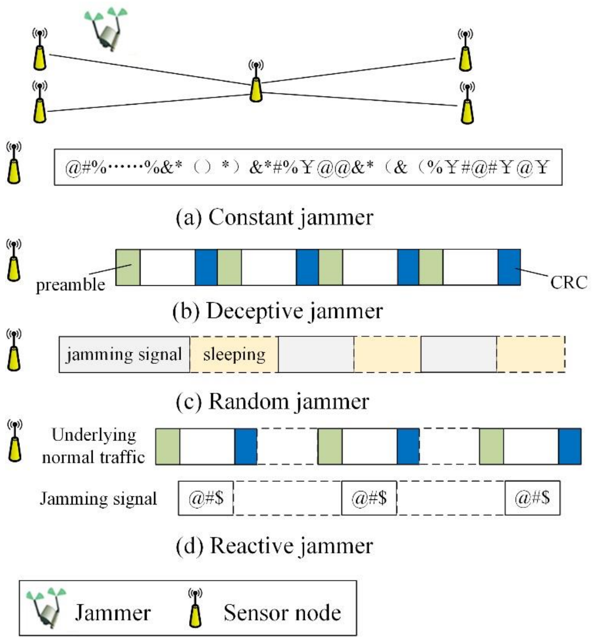
As sensors can be deployed in harsh environment without physical protection, nodes can be easily captured and then compromised [88]. This attack is related to controlling a legitimate node by an adversary and making it a malicious.

5.3.1.3 Replication Node Attack

A compromised node can be easily replicated into a number of clones [89]. A replicated sensor node can be connected in the WSN with same legitimate ID and keys from the original legitimate node. The replicated node can take part in network operations as a normal node without possibility to distinguish between the original legitimate node and its replica. Thus, it is hard to detect this attack based on legitimate IDs and keys. Replicated nodes may be deployed everywhere in the sensor network and later initiate insider attacks [90] in the wireless senor network.

5.3.1.4 Jamming Attack

Jamming attacks consist of keeping communication layer busy and thus prevents a node or multiple nodes from receiving messages. The attack exploits electromagnetic signal to affect or interrupt communications among legitimate nodes. In [91], authors classified jamming attacks into four taxonomies which are illustrated in Figure 2.13.



(a) Constant jammer

(b) Deceptive jammer

(c) Random jammer

(d) Reactive jammer

Figure 2.13 Jamming attack models. Source[29]

Constant jammer continuously emits signals to keep communication channels busy and inhibits legitimate nodes from sending messages. Deceptive jammer emits packets in a random time intervals to deceive sensor nodes from sending messages by keeping them in receiving state. Random jammer changes between jamming and sleeping mainly to save energy. Reactive jammer monitors the communication channels and remain quiet as long as there is no active communication and starts to emit signal as soon as it detects any channel activities.

5.3.2 Attacks in Data Link Layer

5.3.2.1 Intelligent Jamming Attack

Intelligent jamming attack in data link layer is targeting known protocol rules, which can obstruct communications and consume sensor nodes energy. The attacker emits data packets that directly target link layer protocol rules. Constant Jamming is not considered as intelligent attack and the intelligent jamming is limited to deceptive jammer, random jammer and reactive jammer.

[29] Data Collection for Security Measurement in Wireless Sensor Networks: A Survey, IEEE Internet of Things Journal (Volume: 6, Issue: 2, April 2019)

5.3.2.2 Collision Attack

In this attack, the objective is to keep the sensors busy in active communication via several packet retransmission and reception which consume a lot of energy and prevents nodes from communication. The attacker distort byte values of packet before sending them. At the destination node, and because of checksum mismatch, the packet is dropped and receiver node asks for retransmission. Moreover, repeated packet retransmission may consume a lot of resources

5.3.2.3 Denial-of-Sleep Attack

Sleep mode is the mode where sensors saves energy while not communicating[92]. This attack is about preventing a sensor node from falling into a sleep mode in order to exhaust its energy as quickly as possible like in jamming and flooding attacks. This leads to reduce the lifetime of sensor nodes and the wireless sensor network

5.3.3 Attacks in Network Layer

5.3.3.1 Selective forwarding attack

Commonly known as the Gray-hole attack. It is an extension of the Black hole attack. In this attack, the intruder prevents the transmission of certain packets. These will subsequently be deleted by this malicious node. As a result, the network is prevented from functioning correctly and from performing its functions[93]. It should be noted that the choice of packages is based on certain criteria such as: the contents of the packets, source address of the sender, or in a way randomly.

5.3.3.2 Hello flooding attack

In WSN, each node sends a `Hello 'message over a distance of one hop to find neighboring nodes and inform them of its existence. Any node receiving this message knows that it is in the coverage area of the transmitter node. However, during flooding attacks, the malicious node sends a `Hello 'message to the network using a large transmit energy. Therefore, all nodes which receive the message will try to transmit their packets through the malicious node unintentionally. The aim of this attack is to consume the energy of the nodes and prevent their messages from being exchanged which leads to an interruption of the entire network[94] .

5.3.3.3 Wormhole attack

This attack is considered to be one of the most serious against WSNs, it is performed using more than one malicious node. In this attack, a malicious node presents itself as having the shortest routing path to the base station. All the other nodes therefore select it as sender of their packets, resulting in all traffic routed through the compromised node. The attacker tunneled packets from the compromised network node to another malicious node at the other end, which can cause problems in the algorithm routing [95] [96]. The compromised node can also corrupt data packets.

5.3.3.4 Black hole Attack

The Black hole Attack is a special case of elective transfer when a malicious node receives a route request message (RREQ), it announces itself as having the shortest path to the base station without checking its routing during the path finding process. When the source selects the path including the attacking node, all the traffic is routed to the adversary node and this node refuses to redirect data packets and starts dropping them. Instead of forwarding packets selectively as in normal routing process, the attacker will ignore them and source packet will not reaches the destination[97] [98] .

This attack also results in excessive consumption of resources at the sensor nodes level[99].

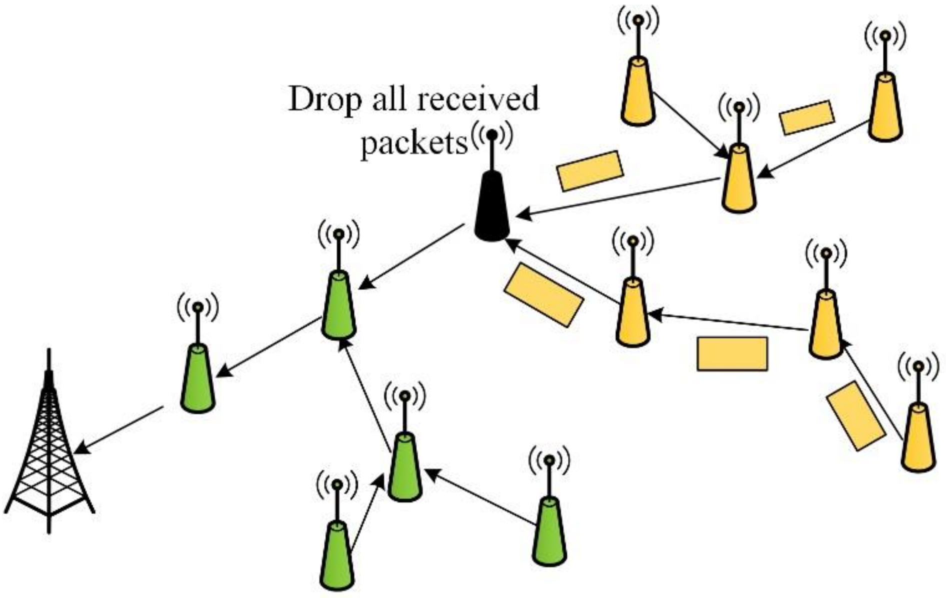


Figure 2.14 Blackhole Attack. Source[30]

[30] Data Collection for Security Measurement in Wireless Sensor Networks: A Survey, IEEE Internet of Things Journal (Volume: 6, Issue: 2, April 2019)

5.3.3.5 Sybil attack:

In WSNs, each node must have only one identity. In Sybil attack, the malicious node creates several fake identities related to other legitimate nodes of the network (by theft or by creating fake ones) [100]. Other nodes can think that this node is ubiquitous and the probability of routing packets from the victim nodes towards the attacking node increases. This attack can degrade the efficiency of several features such as data distribution, data aggregation or filling the list of neighbors or neighboring nodes with non-existent nodes. The goal of the Sybil attack is to change the integrity of the data and routing mechanisms, and can open the door to other cyber-attacks such as selective forwarding, black holes and gray holes. Sybil attack can lead to denial of services.

5.3.3.6 Sinkhole attack:

In this attack, a malicious node convinces its neighbors that it is the closest node to the base station using high transmit power to route all network traffic through itself. Therefore, all packages received will be modified and transmitted to the base station in order to prevent the BS from getting complete and accurate data. This attack can also be carried out by placing the malicious node at the position closest to the base station by the attackers[101] .

5.3.3.7 Clone attack

In this attack, the attacker first starts by capturing the legitimate sensor nodes from the wireless sensor network. Then the attacker gathers all legitimate nodes information, copies them the victim nodes memory to other sensor nodes to create clone nodes. Last step is deploying clone nodes in the network. Once a node is clone, adversary can then launch any other attacks.

5.3.3.8 Manipulating routing information attack

The attacker, aims to change the routing information between two sensor nodes. This attack can be achieved through replaying or spoofing the routing information. In general, the attacker creates routing loops, attracts or repeal network traffic, and extends or shorten source routes. This is an easy passive attack, which is easy to launch and difficult to detect.

5.3.4 Attacks in Transport Layer

5.3.4.1 De-synchronization attack

De-synchronization attack is quite similar to collision attack in data link layer as leading to receiver asking for packet retransmission. In this attack, the attacker changes packets or creates packets with wrong control flags or skipping sequence numbers. The receiver sensor node detects a missing packet and it will request the sender node to retransmit it. When this process is repeated between attacker source and victim destination node, it will exhaust the energy of the receiver.

5.3.4.2 Flooding attack

Flooding attack is a Denial-of-Service (DoS) attack where the attacker sends a large number of packets to a legitimate sensor node. The victim node is kept busy receiving useless packet and consume its power in communications leading to degradation of network lifetime. TPC SYN flooding attack is an example where the attacker sends a high number of packets to request connection establishment with the victim sensor. Once the receiver gets the connection requests, it will reply by sending back an acknowledgement packets and wait for connection establishment. The victim also allocates dedicated memory for transmission control. This attack consume sensor nodes resources and degrade the network performance and may lead to DoS.

5.3.5 Attacks in Application Layer

5.3.5.1 Malicious Code attack

Malicious code attack is a type of application layer attack where an attacker injects a malicious code in a victim node to gain complete control of the node. Once the attacker takes control of the node, other type of attacks can be launched which affect the security of the data transmitted across the network and prevents the network form performing its intended functions.

5.3.5.2 Attack on reliability

Attack on reliability is an attack where the adversary inserts malicious nodes inside the wireless sensor network. These malicious nodes will forges false data or queries and send them as if they are coming from other legitimate nodes. So, the base station will receive false data and at the network level, this attack increase energy consumption and reduce network performance.

6. WSN key disciplines review and discussion

Sensors are tiny devices with limited power (generally a battery for the full sensor lifetime) and reduced computing resources and memory. WSN are deployed for many applications like environment monitoring and health monitoring. In order to reduce the energy consumption and increase the WSN life, adapted network protocols like routing and data aggregation are used. The scheduling of the communications also allow better management of active/idle status and hence optimize the energy consumption. Classical data compression and decompression algorithms are not adapted to WSN as requiring high memory and computation resources. Compressed sensing is an emerging method with practical results in fields like medical imaging. The adoption of compressed sensing in WSN is attracting many research efforts[102] [103] [104] [105]. Ensuring data security is a key challenge for sensitive applications like health and military. Because of the harsh environment where WSNs are deployed and the nature of wireless communication, WSNs is facing several attacks affecting network and the application it serve.

Numerous technologies and disciplines are contributing to manage wireless sensor network constraints. The disciplines that have been reviewed in this chapter are opening the door for various applications with different requirements and approaches.

In general, hierarchical routing protocols have been designed to solve scalability issue, which is the main limitation of most of flat routing protocols. These flat routing protocols with unnecessary overheads lead to an overuse of scarce available bandwidth and thus poor quality of service will be delivered to user's application such high latency and high packet loss ratio.
Cluster formation and maintenance process have been reviewed and a specific focus was done on LEACH protocol. This is a hierarchical protocol that ensure cluster head role is taken by all nodes in a row before a node is elected cluster head for second time and process is repeated.
LEACH protocol optimize energy consumption in the wireless sensor network and enable data aggregation at cluster heads because of its hierarchical structure. Once the Clusters are formed and steady phase start, the network operations begin. Data aggregation is one critical step to deliver sensor readings to the base station. Cluster head nodes consolidate data from cluster members and hence they have a major role in routing data to the base station.

Compressive sensing distribute the compression effort across the sensors in the network. The aggregation function of compressive sensing is designed to compress the data during the sampling which has the advantage of reducing the volume of transmitted data. The compression load is low and the data recovery which require high computation effort is done on the sink side where there is no resource constraints. Compression matrix which satisfies Restricted Isometric Property (RIP) have been reviewed which guarantee the data recovery via research of minimum distance by solving the optimization problem of seeking the sparsest signal.

Many attacks which are targeting wireless sensor network are focusing on data aggregation phase and cluster heads. Energy consumption is highest during packet reception and thus most of the attacks intended to affect network performance or even Deny of Service attacks tries to keep sensor nodes in receiving mode. An attacker would start with a passive attack by monitoring the network and learning or even reading full data if not encrypted. Once the attacker get enough data or if a physical node is compromised, then the adversary node can launch active attacks which can disturb the network functioning and its ability to deliver the service as required by the application it was deployed for.

Data protection is a key topic especially in military and critical applications where lives and high economic interests are relying on the network ability to protect data privacy while it’s being processed or transferred across the network. In the next two chapters, the focus will be put on specific threats and attacks where the proposed techniques allows to secure data privacy while managing energy consumption. Attack detection will be proposed in the last chapter based on machine learning which has the advantage of high detection rate in a wide range of attacks.

7. Conclusion

Security of nodes in WSNs has substantial negative impact on power consumption and thus network performances and lifetime. This influence is essentially revealed by the increase of network overhead and traffic control messages. On the other hand, latency, path throughput and packet delivery ratio show significant decrease at the point where applications might not work properly.

The present chapter was devoted to study the different technologies required to reduce energy impact of node’s security on network performance and assess the efficiency of aggregation and routing protocols in terms of routing overhead, bandwidth consumption and link failure recovery delay. First, we reviewed a taxonomy of data aggregation techniques to understand different strategies that could be flowed to develop efficient routing protocol according to specific network environment. Afterwards, we presented compressive sensing and how it can be applied to collecting and compressing data at nodes level with a low computational overhead for matrix multiplications. Data recovery is possible because of the sparse signal of sensed data. Sparse data transformation have been viewed via projection into a sparse signal and the choice of sensing matrix. The wireless nature of WSNs make them vulnerable to wireless attacks and the energy constraints increase the ability of attackers to degrade the network performance and data protection. Various attacks have been presented depending on which communication layer they target. This study shows that the combination of compressive sensing with hierarchical routing protocols have the potential to deliver good performances in terms of energy consumption even for large wireless sensor networks. We concluded that security is possible to manage when we decrease communication and amount of data. However, secure operations should be optimized to fit with nowadays wireless sensor network application and security requirements.

Chapter 3

Novel secure data collection schemes based on homomorphic encryption and compressive sensing

1. Introduction

Since last decades, we have been assisting to the rapid development of shared applications that require networking facilities to deliver virtual online services anywhere at any time, such as remote storing and processing used in cloud services and remote sensing used in internet of things applications. In this context, Wireless sensor networks have been massively used in wide applications' areas, particularly in controlling and monitoring environment parameters, weapon control and tracking, urban intelligent transportation, smart cities and other fields [8].

Mobile wireless networks are cost effective, dynamic self-organized and temporary network system. They are built in the area where there is no pre-established infrastructure like battlefield or to address an urgent networking need like rescue operation.

WSN is a set of autonomous devices that are able to set up networking facilities without the support of any existing infrastructure or any central administration. Sensor nodes in WSNs act as host and router at the same time. Thus they are responsible of data acquisition, processing and storing at application layer and data flows management like routing and forwarding at network layer [8]. However, wireless sensor nodes present some intrinsic weaknesses, which affect their performances while involved in the global framework service delivery process. Such limitations could be summarized in three main limitation: Power and processing limitations and security [106].

In the following section we present an overview on the importance of Compressive Sensing with Cryptosystem based Techniques in WSN. In section 3 related works are presented to explain how research community is handling secure data aggregation in WSN. A set of previous works, which studied the behavior of different secure routing protocols in WSNs based on multiple criteria such as node density, energy, propagation models, and attack detection. Section 4 is dedicated to present the system model and problem formulation. The "Cluster-based Secure Data Aggregation" (CSDA) scheme is reviewed as offering high performance and would serve as reference for comparing our two new schemes. In section 5, we review Homomorphic encryption and Paillier Cryptosystem based Key Distribution and Management. Also we present Secure Data Transmission and Attack Detection in WSN before focusing on the new "Paillier Cryptosystem and Compressive Sensing based Routing" (PC2SR) protocol. Security and Cost analysis for the proposed protocol PC2SR will follow. The second suggested protocol "Cluster-based Semi-Homomorphic Encryption Aggregated Data" (CSHEAD) is presented in section 6. In section 7, we present some well-known simulation platform to motivate the choice of network simulator NS2 to carry out our study. In section 8, we present the four conventional performance metrics namely latency or delay, PDR, Throughput and routing overhead and then we add attack detection and energy consumption to enhance the accuracy of performance validation framework. Finally simulation results are presented and discussed. We conclude in section 9 by reminding key results and improvements of the suggested schemes.

2. Compressive Sensing with Cryptosystem based Techniques

Wireless Sensor Networks (WSNs) comprise battery-powered, low-cost tiny devices to monitor an unattended area [107] [108]. Due to the flexible characteristics, WSNs are promising future in numerous applications like surveillance systems, construction fields, and medical automation [109] [110]. However, data collection in WSN is a tedious task due to massive-scale sensor deployment. Large scale sensor devices institute a vast amount of data, and thus, it increases the transmission cost of the network. It is merely insupportable to transmit the entire data to the base station due to constrained energy and bandwidth of sensor devices. To alleviate such an issue, the WSN necessitates clustering algorithms to diminish the transmission cost by virtually exploiting the distributive sensor characteristics [111]. Compressive sensing-based data collection is a popular technique to attain a better tradeoff between reliability and energy consumption with lower transmission costs [112]. The compressive sensing model can surpass the pitfalls of traditional data collection methods of WSN by exploring the compressible signal spares [113]. However, the compressive sensing model increases the sampling rate under emergencies, and hence, it is crucial to reduce the sampling rate without degrading the network efficiency. Some current works employ spatial or temporal relationships to minimize the communication cost of correlated data [114]. However, such models accomplish limited efficacy due to the data collection over a long period. Therefore, it is essential to consider Spatio-temporal correlations in compressive sensing to be performed in a cluster and employ an efficient cryptosystem from CH to base station communication to maximize network efficiency. Albeit, the compressed sensing data collection model, is suffered by some security threats due to the deployment complexities of WSN.

The WSN is vulnerable to various security threats due to the loosely coupled unreliable wireless links [115]. To reduce network efficiency, an attacker may attempt to inject false data into the network by altering the original data. Such type of attack is named as false data injection attack. Encrypting the messages is a fundamental method to ensure security and to detect attacks in wireless networks. According to the encryption type, the traditional secure routing protocols are classified into hop-by-hop and end-to-end. The hop-by-hop encryption method lack of offering high security, as it necessitates to share the keys with intermediate nodes for decryption, resulting in false data injected into the network. The end-to-end encryption method overcomes the issues of hop-by-hop by enabling the intermediate nodes to aggregate the received data without performing decryption straightly. In end-to-end encryption, the base station involves decrypting the messages, and there is no chance to get the real data by an attacker. Homomorphic cryptography is a usual method for providing end-to-end encryption in WSN [116] [117]. However, the homomorphic encryption method only deals with simple arithmetic operations on cipher texts, and it introduces some noise in the decryption of data. Thus, it diminishes the decryption efficiency. Therefore, it is essential to propose a novel compressive aggregation model with high security for achieving better efficiency in WSN.

This chapter proposes two novel lightweight, secure data aggregation routing protocol, named as “Paillier Cryptosystem and Compressive Sensing based Routing” (PC2SR) protocol and “Cluster-based Semi-Homomorphic Encryption Aggregated Data” (CSHEAD). The primary objective of PC2SR and CSHEAD is to assure high security against false data injection attacks and reduce transmission cost over energy-constrained WSN. The significant contributions of the proposed PC2SR and CSHEAD are as follows.

- Effectively reduce the communication cost and improve the WSN security level in terms of data authentication, confidentiality, and integrity using lightweight mechanisms. By combining the paillier cryptosystem with compressive sensing, the proposed PC2SR and CSHEAD can reduce communication costs and detect various security threats over WSN.

- Initially, the provision of short-term paillier security keys for each node ensures data authentication and also solves the security issues in secure data gathering. The lightweight key refreshing mechanism enables the nodes to update the keys itself over a particular period and prevents the keys from inferring by an attacker node.
- With aiming to prolong the network lifetime, the PC2SR is implemented over a clustered WSN as well as CSHEAD. A compressive sensing based data gathering model based on the Spatio-temporal sparsity measurement matrix within intra-cluster significantly reduces the communication cost and also improves the data collection efficiency.
- In PC2SR, the CH nodes involve in aggregating the compressive data and routes the data through the inter-cluster tree to the base station. Integration of zero noise factor with transmitted data assists PC2SR to detect and alleviate the malicious behaviors successfully. Thus, the PC2SR ensures high security in terms of integrity and confidentiality without degrading routing performance.
- Finally, the effectiveness of the proposed PC2SR and CSHEAD is evaluated using Network Simulator-2 (NS2). The efficacy of PC2SR and CSHEAD is estimated using some performance metrics that are detection accuracy, overhead, packet delivery ratio, delay, and energy consumption.

3. Related work

Compressive sensing is an efficient paradigm to aggregate the data in WSNs. The works in [118] and [119] survey and discuss the recent compressive sensing based data collection methods proposed for WSNs. For a comprehensive study, the existing works are divided into two types that are compressive sensing based and compressive sensing with cryptosystem based.

To reduce the computational cost and to enhance the efficiency of privacy, a secure data collection scheme based on compressive sensing (SeDC) has been presented in [120]. The SeDC analyzes the security issues associated with compressive sensing based data collection over WSN. It also defines two various attack models. Further, the SeDC utilizes an asymmetric semi-homomorphic cryptography mechanism for encryption. Thus, the SeDC rectifies the difficulties in secret key issuing and management processes. The SeDC is mainly fit for applications that require high security. The work in [121] introduces a Privacy Preserving Compressive Sensing (PPCS) technique that utilizes a homomorphic obfuscation property to perform compressive sensing. The PPCS recovers the trajectories based on the crowd sensing method and also preserves privacy. Toward the intention of offering security and magnifying the data collection efficiency of WSN, an adaptable, secure compressive sensing–based data collection scheme has been introduced in [122] for distributed WSNs. The flexible scheme incorporates the public key cryptosystem with compressive sensing to solve the key dissemination and communication cost issues. In [123], a combination of compressed sensing and homomorphic encryption method is utilized to minimize the communication load and to perform encryption operations based on arithmetic functions, respectively. Such a technique also alleviates the decryption operation at every node, and it enables the base station to perform and handle the decryption complexities. In realistic WSN, the base stations have more resources, and it computes the complex arithmetic operations of data decryption. Thus, the compressing sensing with a homomorphic cryptography mechanism assures a high privacy level and also assists deployment of WSN in highly sensitive applications like surveillance system and healthcare automation. A novel Multi-functiOnal secure Data Aggregation scheme (MODA) has been presented in [124]. The MODA performs encoding of the raw data and provides building blocks to accomplish data fusion based on multi-functional. The MODA incorporates a homomorphic cryptosystem to aggregate the cipher text, and it assures end to end security in the network. In[125], authors presented Multi-Functional and Multi-Dimensional Secure Data Aggregation Schemes in WSNs. The scheme is designed to support applications need in term of heterogeneous sensed data requiring multi-dimensional data aggregation and multi-functional data analysis. Moreover, two supportive solutions that are RandOm selected encryption-based Data Aggregation (RODA) and COmpression based Data Aggregation (CODA) are introduced according to MODA to reduce the communication cost and energy depletion.

CSDA [126] is an energy efficient secure data aggregation scheme based on cluster privacy preserving. The scheme is classified into three steps. The first step is cluster formation and second step involves data aggregation in one cluster using slice assemble technology. The third step is the data aggregation between clusters. In the first two steps of CSDA, the cluster head node is used to data aggregation, while other member nodes are responsible for keeping watch on head node’s operations. Furthermore, a few conventional routing protocols consolidate the cryptography mechanism with compressive sensing to enhance the security and efficiency of WSN. However, such methods still face some imperfections in WSN security requirements like data integrity and confidentiality and also lack in utilizing the network resources effectively.

4 System Model

Compressive sensing is a novel method that accomplishes a projective transformation of sparse measurements of signals in high-dimensional space. The proposed PC2SR exploits the advantages of compressive sensing to prolong the network lifetime and to optimize network efficiency. The network is considered as a communication graph G= (N, E), where N represents the number of WSN nodes, and E refers to the direct communication links among two entities. The term N comprises both sensor nodes and the base station. The PC2SR considers a multi-hop wireless sensor network for efficient data gathering, which consists of S sensor nodes S = (1,2,3 …, s) and one base station (BS). The PC2SR deploys the sensor nodes randomly in a specific square network region. To reduce the energy expenditure, the PC2SR divides the S into clusters, C=(c_1, c_2, …..c_S). Each C has a cluster head to aggregate the compressed data of corresponding C and forwards it to the BS over a specific time interval T=(t_1, t_2,…t_i). Further, the sensor nodes periodically sample spatial-temporal data from a corresponding monitoring area, and the selected compressed nodes perform compressive sensing within C. For instance, the ith sensor node measures m readings for every T, and the measurement reading is denoted as a signal vector $\mathbf{x}_i = |(x_{1,}\ x_2, \ldots, x_i)|_T$. The generated sensor readings are compressively sensed among CM nodes and forwards to the CH node for data aggregation and forwarding. The homomorphic encryption method named as paillier cryptosystem is used to assure security in PC2SR. For secure communication, every node has a public and private key pair (K_{Pub}, K_{pri}) offered using the paillier key generation model. Moreover, the PC2SR is a lightweight security model that preserves data integrity and data confidentiality of the WSN.

4.1 Threat Model

In a false data injection attack, an attacker injects false measurements into the network with aiming to disrupt the routing performance. The false injection attacker affects both the integrity and confidentiality level of data generated by sensor nodes. Generally, the sensor devices are deployed in an unattended area, and a malicious node easily compromises the sensor nodes or snoop the security keys of such nodes. The main intention of a false injection attacker is to alter the data in the packets or drop and inject false measurement information during data routing. The false data injection is done through compromised sensor nodes, and it is crucial to detect such attacks quickly for preventing the devastation of valuable network resources such as energy and bandwidth. In PC2SR, the attacker sensor node (S_{Mal}) should have high energy, memory, and processing capabilities than the other legitimate sensor nodes (S_{Leg}) in PC2SR. The attacker S_{Mal} injects false measurements D_{bad} into the original measures (D_{ori}), and the malicious measurement report, $R_{Mal} = D_{bad} + D_{ori}$. The PC2SR uses a zero noise factor with original measurements for detecting and isolating the malicious nodes from the network.

4.2 Problem Formulation

Compressive sensing based data aggregation is a proliferated technique that highly suitable for resource-constrained networks, as it reduces the communication cost-efficiently. Due to the dubious wireless medium, WSN is vulnerable to security attacks; especially false data injection attack is a severe issue. For that, a few existing works exploit the cryptography mechanism with compressive sensing for maximizing the system security and efficiency. However, the conventional security methods face some imperfections in terms of integrity and confidentiality in secure data aggregation and forwarding over WSN. Initially, the homomorphic cryptography necessitates the plaintext of sensor measurements as a finite set like $Ʈ_{ij}S_j \in R$ but is not a finite set in real-time. Thus, it reduces the encryption efficiency of cryptography mechanisms. For providing high security and efficiency, the PC2SR employs a homomorphic security mechanism that includes pair of keys (K_{Pub}, K_{pri}) generated using a paillier key generation algorithm to the WSN nodes for secure encryption and decryption. To efficiently utilize the restrained energy resources, the PC2SR implements its protocol design over a clustered architecture. In PC2SR, the role of CH is periodically rotated among the clusters, and there is a chance for an attacker to obtain the security keys quickly and retrieves the original measurements to launch an attack into the network. For instance, a set of sensor nodes S = (1,2,3 …, s) generates original data (D_{ori}) over a specific period T. It forwards the data to the CH using encryption keys (Kpub, Kpri)$_{CM}$. Further, the CH aggregates the received data by decrypting its security keys (Kpub, Kpri)$_{CH}$. An attacker node easily compromises the sensor node CM and CH due to the vulnerable WSN characteristics. It tries to manipulate the D_{ori} or inject false information with D_{ori} (D_{ori}+D_{bad}) to reduce system efficacy. To rectify such an issue, the PC2SR incorporates an efficient, lightweight key generation and refreshing mechanism using a paillier cryptosystem with compressive sensing.

4.3 CSDA aggregation scheme

CSDA is based on the data slicing mechanism combined with encryption to protect the data. Slice assembly technology means each node divides its data into pieces and send each piece to the other cluster members. Let's assume the cluster has C_i members. Each node will divide its data into M_i pieces and send the $M_i - 1$ to the cluster members. Each sensor node establish a secure link with its neighbors which share the same encryption key (ki).

The example below shows three neighbor nodes A, B, and C dividing their private data a, b, and c into three slices separately: a1, a2, a3, b1, b2, b3 and c1, c2, c3. They deliver encrypted values to each other and keep the 3rd piece for themselves.

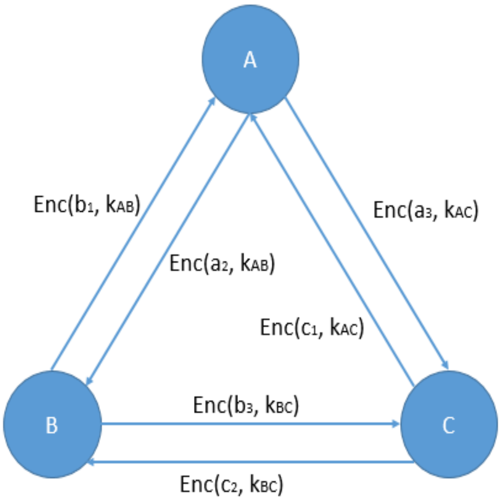


Figure 3.1 Data exchanging : Encrypt and deliver part of values

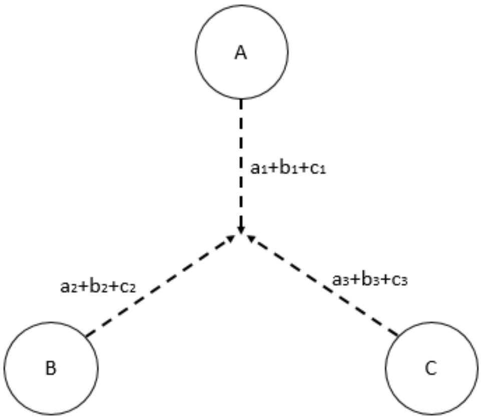


Figure 3.2 Broadcast the result of aggregation

A receive b1 and c1 and have a1 which was not shared. A will calculate the value of a1 + b1 + c1. Same do node B to calculate the value of a2 + b2 + c2, and a3 + b3 + c3 is to be calculated by node C. Figure 3.2 shows the nodes broadcasting the three values, and the three nodes add up three values to infer the values of a + b + c.

In CSDA scheme, hop-by-hop encrypted data aggregation is used, each non-Leaf node is decrypting the received data. This scenario present a high risk for data confidentiality if an attacker compromises a node. The attacker can then get access to the encryption key and become able to perform the decryption process.

4.4 Homomorphic Encryption

Homomorphic encryption [127] allow arithmetic operations on cipher texts without losing capability to recover transformed data. Semi-homomorphic cryptosystems are asymmetric homomorphic [128] encryption algorithm with public and private keys. Let's denote E the encryption function and m_{i} the message or private data of node i. In Fig. 3, the aggregation of encrypted data transform the multiplication of 2 encrypted messages m_1 and m_2 into addition.

$$\boldsymbol{E(m_1).E(m_2) = E(m_1 + \ m_2)} \qquad \textbf{(1)}$$

Sensor node S1 encrypt its data m1 and send encrypted message E(m1) to the parent node S3. Sensor node S2 encrypt its reading m2 and encrypt it before sending E(m2) to S3.

Aggregation applied at parent node S3 is performed by multiplying the received encrypted messages from S1 and S2. Sensor S3 encrypt its own reading m3 and multiply E(m3) with received encrypted messages E(m1) and E(m2) before forwarding the result $E(m_1).E(m_2).E(m_3)$ to the Sink.

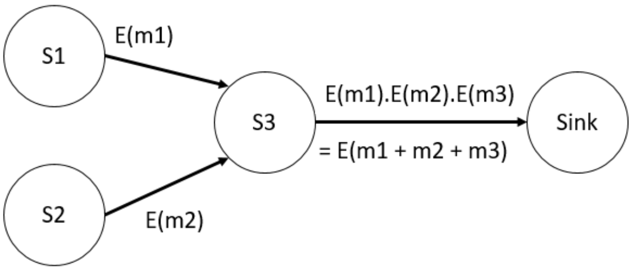


Figure 3.3 Semi-homomorphic encryption transform multiplication of encrypted data into an encrypted message of the additive result from initial messages.

Given a scalar t, then:

$$\boldsymbol{E(t.m) = E(\textstyle\sum_{i=1}^{t} m) = \prod_{i=1}^{t} E(m) = \ E^{t}\,(m)} \qquad \textbf{(2)}$$

$$\textbf{And} \quad \boldsymbol{E(\textstyle\sum_i t_i.m_i) = \prod_i E(t_i.m_i) = \ \prod_i E^{t_i}(m_i)} \qquad \textbf{(3)}$$

At the sink level and by applying (1), the received message is same as $E(m_1 + \ m_2 + m_3)$ and hence after decryption sink will get the sum value $m_1 + \ m_2 + m_3$.

The sensed data collected by the cluster members is encrypted using public key based on semi-homomorphic encryption. End-to-end encrypted data aggregation uses homomorphic encryption to apply certain aggregation functions such as addition or multiplication on the encrypted data. No decryption is required during the data routing from sensor node until delivery to the sink. Therefore, this scenario reduces the decryption workload in the network. In addition, in case of sensor node physically compromised, the data confidentiality during transmission is not affected, as the decryption key is not available on the network. The nodes have the public key for encryption but the decryption key is known by the sink only. In this chapter, Benaloh and Paillier cryptosystem have been used for sensor reading confidentiality protection.

5. PC2SR Protocol Design

The compressive sensing based data collection method is an emergent paradigm for diverse application fields of WSNs, which consolidates data acquisition and compression by employing compressible signals. To reduce communication cost and energy depletion, the compressive sensing model retrieves the original measurement data from a limited number of sensor nodes that perform compressive sensing, instead of collecting total measurements of WSN. In WSN, the sensor readings are mostly spatial and temporal dependence in different applications, and it is essential to compress the data in both directions for attaining better network efficiency. For that, the proposed PC2SR model employs the Spatio-temporal sparsity measurement matrix over a clustered WSN. Further, the compressed data at cluster heads is forwarded to the base station in a single and multi-hop manner. Data forwarding without performing encryption and decryption leads to several security threats, and thus, it reduces the WSN efficiency, resulting in insecure environmental monitoring. To alleviate the malicious behavior and to improve the WSN security, the PC2SR employs a semi-homomorphic cryptosystem, named as paillier for encryption and decryption of compressed sensing data. Instead of offering a long term public/private key pair among two entities, the paillier cryptosystem in PC2SR updates the keys over a specific time interval based on a lightweight key updating mechanism. For efficient attack detection, the PC2SR integrates a zero noise factor with aggregated data before data forwarding. Using the additional zero noise factor and modified paillier cryptosystem, the PC2SR efficiently satisfies the WSN security in terms of integrity and confidentiality. Moreover, the PC2SR reduces the communication cost and resource usage of WSNs without compromising the data security level by utilizing a practical secure paillier cryptography based compressive sensing model. Figure 1 shows the block diagram of PC2SR. To intelligibly explain the protocol process, the proposed PC2SR designs three mechanisms that are Paillier cryptosystem based Key distribution and management, Intra-cluster Data Gathering, and Secure Data Transmission.

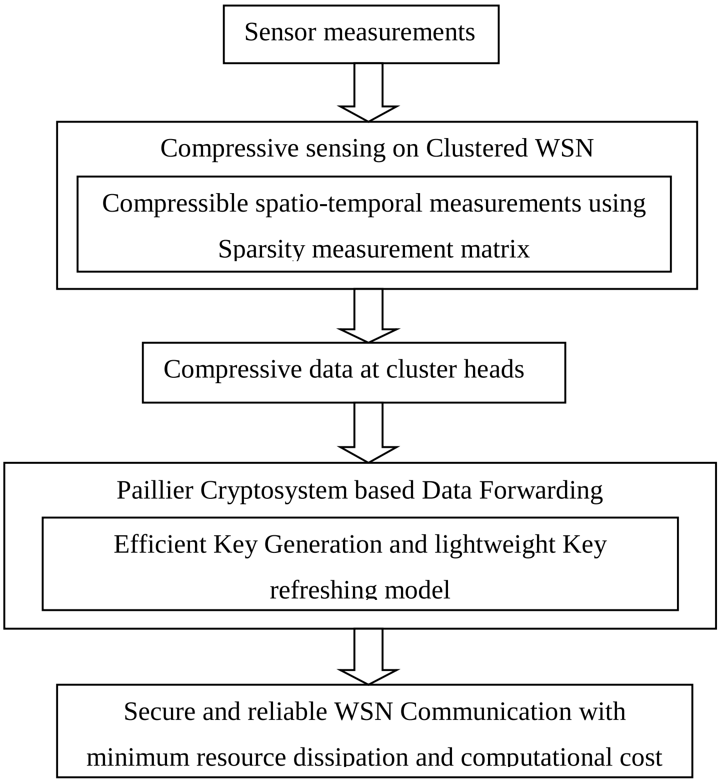

Sensor measurements
Compressive sensing on Clustered WSN
Compressible spatio-temporal measurements using Sparsity measurement matrix
Compressive data at cluster heads
Paillier Cryptosystem based Data Forwarding
Efficient Key Generation and lightweight Key refreshing model
Secure and reliable WSN Communication with minimum resource dissipation and computational cost

Figure 3.4 Block Diagram of PC2SR

5.1 Paillier Cryptosystem based Key Distribution and Management

The proposed PC2SR employs a modified paillier cryptosystem to distribute and manage the security keys among WSN nodes. Paillier cryptosystem is a chipper based homomorphic model that enables two types of keys that are public and private for ensuring rich security in the network. In PC2SR, each sensor node has a public-private key pair for encrypting the data.

The public key is widely distributed among the nodes, whereas the weak private key is maintained secretly. However, an ingenious attack can compromise and obtains the original data for introducing malicious behavior into the network. To rectify such an issue, the Paillier cryptosystem model offers a lightweight key refreshing mechanism to update the keys over a specific time interval. The Paillier cryptosystem preserves data privacy and neglects node compromise by endowing the additive operations on the encrypted multiplicative cipher text.

Key Generation Process: In PC2SR, two types of keys, such as public (K_{pub}) and private (K_{pri}), are generated. Steps to generate Paillier keys are described as follows. The public key K_{pub} = (N, g, h = g_ mod N) in which g is a base, and it is estimated using two prime numbers p and q (p=2/p+1 and q=2/q+1). Consequently, the private key $K_{pri} \in (1, N^2/2)$ is generated. In PC2SR, each sensor node individually has a key pair (K_{pub}, K_{pri}) for intra and inter-cluster data encryption and decryption. The entire WSN nodes know the $K_{pub,}$ and the nodes hide their private key K_{pri}. However, an attacker may compromise a node and get the K_{pri} for retrieving the original measurement data by decrypting the encrypted message using K_{pri}. Such an issue can be avoided in PC2SR due by updating the K_{pri} using the factor γ. Moreover, the attacker has to get recent K_{pri} generated using υ for injecting malicious behavior, and it is challenging in PC2SR due to the key updating process. The paillier cryptosystem performs additive operations on the aggregated data created using the (K_{pub}, K_{pri}). Further, the BS can decrypt the cipher text using its K_{pri}. The BS also knows K_{pri} of the sensor nodes in the network. Compared to the other traditional cryptosystem, the paillier cryptosystem attains better performance in the network in which multiple nodes can send the data to a single base station.

Steps for key generation:

- Select two random prime numbers p and q, which are independent of each other.
- Compute gcd(pq, (p-1), (q-1)). This property assures that the selected prime number's length is equal.
- Calculate N=pq. Select two random integers g=$Z_{n^2}{}^*$.
- Assures that the term N is divided by order of g.
- The public key is K_{pub} = (N, g, h = g_ mod N).
- The private key is $K_{pri} \in (1, N^2/2)$.
- Updating the secret key K_{pri} for a time interval T using γ.

5.2 Intra-cluster Data Aggregation based on Compressive Sensing

To efficiently handle the network resources, the proposed PC2SR divides the network nodes into clusters. Each cluster C has several CM nodes and also a CH node. Among the CM nodes, the PC2SR randomly selects many compressive nodes in each time slot T for data aggregation. The WSN nodes are distributed over a specific network area X x Y. The proximity sensor nodes have strong spatial-temporal correlations, and the WSN nodes are divided into C number of clusters based on node location. In PC2SR, each sensor measures its readings over a particular time slot T is expressed as follows.

$$\mathrm{S_n(T) = \{M_1(t_1), M_2(t_2) \ldots \ldots . M_n(t_r)\} \ldots \ldots . (1)}$$

In equation (1), $S_n(T)$ denotes the n_{th} sensor node at time interval T and T=$\{\mathrm{t_1},\ \mathrm{t_2}, \ldots .\ \mathrm{t_r}\}$. Further, the total measurements of n number of CM nodes within cluster C_n, $S_n(T) \in C_s$ is estimated using the following equation.

$$\mathrm{S_n(T)|_{C_s} = \sum_{n=1}^{CM} S_n(T) \ldots \ldots \ldots \ldots \ldots . (2)}$$

Generally, the WSN nodes generate a vast amount of data, and hence, transmitting all data to BS through CH increases the communication cost and also drains the node battery power rapidly, resulting in decreased network lifetime. The compressive sensing based data gathering model of PC2SR significantly reduces the communication cost by randomly selects the Sr number of sensor nodes from each cluster to send the compressive data to the CH nodes. The spatiotemporal sparse measurement matrix denoted as $\Phi(\mathrm{s,t})$over a particular time interval is estimated as follows.

$$[\Phi(\mathrm{s,t})]_\mathrm{T} = \left[\begin{pmatrix} \varphi(\mathrm{s_1t_1}) & \cdots & \varphi(\mathrm{s_rt_1}) \\ \vdots & \ddots & \vdots \\ \varphi(\mathrm{s_1t_r}) & \cdots & \varphi(\mathrm{s_rt_r}) \end{pmatrix}\right]_{\mathrm{s_r}} \ldots \ldots \ldots . (3)$$

In equation (3), each row represents the spatial correlation of the compressible sensor data. For instance, the spatial correlation of sensor data is denoted as $\Phi(S) = (\mathrm{s_1, s_2, \ldots, s_r})$and temporal correlation of sensor data is represented as $\Phi(\mathrm{t}) = (\mathrm{t_1, t_2, \ldots, t_r})$. For every T period, the $S_r \in C_i$ nodes send the compressive data to the CH nodes for data aggregation. Its public key Kpub encrypts the compressive data of an Sr node, and the CH node retrieves the original compressed data using its K_{pri} for data aggregation. Further, the CH node exploits a multiplicative homomorphic property to encrypt the compressive data using its (K_{pub}, K_{pri}). The data aggregation done at CH node is demonstrated in equation (4).

$$\mathrm{S_r(p)} = \left(\mathrm{p_{s_1}} \otimes \mathrm{p_{s_2}} \otimes \ldots . \otimes \mathrm{p_{s_r}}\right)_{\mathrm{K_{pub}}} \ldots \ldots \ldots . (4)$$

$$\mathrm{CH_{agg}\ (C)} = \prod_{\mathrm{s_r}} \mathrm{S_r}(p) \ldots \ldots \ldots \ldots (5)$$

To shield the original data, the PC2SR utilizes a zero noise factor (N_k) to the aggregated data. By decrypting the original data with the noise factor, the BS detects the attacker. Finally, the terms $\mathrm{CH_{agg}}$ and $\mathrm{N_k}$ is estimated using the following equations (6) and (7).

$$\mathrm{CH_{agg}} = \left(\sum_{s_r} \mathrm{CH_{agg}\ (C)}\right)_{\mathrm{K_{pri}}} + (\mathrm{N_k})_{\mathrm{K_{pri}}} \quad \ldots \ldots \ldots . (6)$$

Where,

$$\mathrm{N_k} = \begin{cases} \sum_{\mathrm{i=1}}^{\mathrm{T}} \mathrm{N_{S_1 \in C_i}} = 0 \\ \sum_{\mathrm{i=1}}^{\mathrm{T}} \mathrm{N_{S_2 \in C_i}} = 0 \\ \quad \ldots \ldots , \\ \sum_{\mathrm{i=1}}^{\mathrm{T}} \mathrm{N_{S_r \in C_i}} = 0 \end{cases} \qquad \ldots \ldots \ldots . (7)$$

In equation (6), the term $\mathrm{CH_{agg}(c)}$ represents the multiplicative cipher text generated by the CH node. In equation (6), K_{pri} is estimated over m set of T period denoted as T_m follows.

$$\mathrm{K_{pri}} = \begin{bmatrix} \mathrm{K_2} = \gamma * (\mathrm{K_1}/2) \\ \mathrm{K_1} = \gamma * (\mathrm{K_2}/2) \\ \ldots \ldots . . , \\ \mathrm{K_n} = \gamma * (\mathrm{K_{n-1}}/2) \end{bmatrix}_{\mathrm{T_m}} \qquad \ldots \ldots \ldots \ldots . . (8)$$

Further, the CH performs an additive operation on multiplicative cipher text $\mathrm{CH_{agg}}(c)$ without performing decryption. The final $\mathrm{CH_{agg}}$ is only decrypted using the private key of BS. Similarly, each CH encrypts the aggregated data using a paillier cryptosystem for secure data transmission. By concealing the original data from CH nodes, the proposed PC2SR enhances the integrity and confidentiality of the original readings and also improves the overall network efficacy.

5.3 Secure Data Transmission and Attack Detection

After encryption, the aggregated data of the CH node is forwarded to the base station in a single of multi-hop manner. The CH nodes that are closest to the nodes straightly transmit the data to BS. Otherwise, the CH nodes form the inter-cluster tree structure for secure data forwarding. During data forwarding, the data is additively aggregated at each intermediate CH node without performing decryption. The data aggregation at an intermediate node is depicted in the following equation.

$$\mathrm{ICH_{agg}} = \sum_{\mathrm{C}=1}^{\mathrm{h}-1} \mathrm{CH_{agg}} \ldots\ldots\ldots..(9)$$

$$\mathrm{D_{orig}} = \mathrm{dec}\big(\mathrm{ICH_{agg}}\big)_{\mathrm{K_{pri}}} \ldots\ldots\ldots\ldots(10)$$

In equation (9), the term ICH denotes an intermediate CH node of an inter-cluster aggregation tree. The term h denotes the number of hops, and h-1 denotes the direct hop of a base station. Finally, the BS receives the aggregated cipher text of compressive data, and it decrypts the message using its private key K_{pri}. The BS obtains the original reading $\mathrm{D_{orig}}$ and $\mathrm{N_k}$ values from the decrypted data and verifies the message based on the noise factor $\mathrm{N_k}$. If the noise value is equal to zero, the BS concludes that the original data is not manipulated or falsely injected by an attacker node. Otherwise, it decides that the received data is manipulated or falsely injected by a malicious node. Further, the BS applies the reverse decryption process over the inter-cluster routing tree for detecting the attacker node. Moreover, the PC2SR attains better routing performance with high security and minimum transmission cost. The protocol process of PC2SR is described in algorithm 1.

Algorithm 1 PC2SR Protocol Design

Input: Sensor Measurements, Compressive Matrix, and paillier keys

Output: Secure Data Transmission

```
Each node do {
    Measures the sensor data periodically for a time period T;
            }
Each Cluster Head do {
    Selects $S_r$ nodes randomly from $C_s$ for every T;
    Sr nodes do {
        Constructs $[\Phi(s,t)]_T$ based on compressive sensing;
        Encrypt $[\Phi(s,t)]_T$ using the key $K_{pub}$ of CH;
        Forwards the encrypted compressive data $S_r(p)$ to the CH for aggregation;
                }
        Receives an additive cipher text $S_r(p)$ from Sr;
        Perform multiplicative operation on $S_r(p)$;
        Generate $N_k$ value for $S_r(p)$;
        Forwards the $CH_{agg}$ data to the BS over single or multi-hop;
                }
Base Station do {
            Decrypt the $CH_{agg}$ using its $K_{pri}$;
            Recover compressed data $[\Phi(s,t)]_T$ based on compressive sensing
            Retrieves the original information using its ($K_{pub}$, $K_{pri}$);
                Verifies the $D_{ori}$ using the noise factor;
                If (Nk≠0) {
                Perform reverse decryption process for attack detection;
                    Otherwise {
                        Accepts the $D_{ori}$;
                            }
                        }
            }
```

5.4. Analysis of PC2SR

The proposed PC2SR utilizes a paillier cryptosystem for high security and compressive sensing based data aggregation model to reduce communication cost. The effectiveness and complexity of PC2SR are analyzed in terms of security and cost.

5.4.1 Security Analysis

The proposed PC2SR resists various attacks such as false injection, data modification, and eavesdropping attacks. The proof for attack resistance is described as follows.

False Injection Attack: During the false injection attack, an attacker tries to inject malicious information with the original sensor reading, $R_{Mal} = D_{bad} + D_{ori}$. In PC2SR, an adversary cannot inject malicious data without obtaining the paillier security keys.

Proof: In PC2SR, set of CM nodes (CM=s1,s2, ….sr) of ith cluster Ci forward the cipher text of Dori to the CH nodes for data aggregation. Further, the aggregator CH encrypts data using its paillier keys and forwards it to the BS. Therefore, an attacker cannot obtain the D_{ori} without knowing the paillier keys. Although, there is less chance to an ingenious attacker can inject malicious information with the cipher text of D_{ori} traveled along the route. To rectify such problem, the PC2SR protocol employs a non-zero noise factor, $\mathrm{N_k}$ in data aggregation and forwarding. By verifying the $\mathrm{N_k}$, such malicious activity is detected and isolated at BS.

Data Modification Attack: The malicious node obtains the D_{ori} by spoofing the security keys or compromising the intermediate nodes to get the security keys. In PC2SR, it is very tedious to obtain the paillier security keys due to frequent key updating process, and it is no worthless to compromise the intermediate nodes for retrieving the D_{ori}, as the intermediate nodes do not know about D_{ori}.

Proof: Assume that the data modification attacker in PC2SR modifies the cipher text of D_{ori} along the traveling path. During an attack launching, the attacker should modify the cipher text that comprises original data D_{ori} and noise value $\mathrm{N_k}$. Finally, the BS receives the message and decrypts the encrypted aggregated data $\mathrm{ICH_{agg}}$ using its private key. The BS verifies the value $\mathrm{N_k}$. If the value $\mathrm{N_k} = 0$, the BS recognizes that the data is not modified by an attacker and accepts the D_{ori}. Otherwise, it decides the cipher text of D_{ori} is modified by an attacker. Moreover, the BS performs reverse decryption operation for detecting the malicious node.

Eavesdropping Attack: To reveal the private data of devices, the eavesdropping attacker listens to communication behavior. A malicious node cannot obtain the private data of sensor devices by spying the routing behaviors.

Proof: For assuring secure communication among devices, the PC2SR permits the nodes to append a noise value equal to zero with all original information transmitted to BS. Thus, it prevents an attacker from revealing the original private information. Also, the transmitted information is encrypted using the fresh security keys. So, PC2SR prevents the malicious node from eavesdropping the original information.

5.4.2 Cost Analysis

The proposed PC2SR employs simple computational operations like Kronecker product $\otimes$ and simple multiplicative operations to design the Spatio-temporal measurement matrix and compressive data aggregation. Generally, the WSN nodes are deployed in two or three-dimensional space, and the Spatio-temporal sensor readings obtained from such an environment are not complicated. The multi-dimensional WSN deployment methods also reduce the computational complexity of sparse measurement matrix design. After matrix formation, the PC2SR instructs the nodes to encrypt the data using the paillier cryptosystem before data aggregation and transmission. Assume that S numbers of sensor nodes are deployed in the two-dimensional area (X x Y), and each node has its Spatio-temporal readings $[\Phi(s, t)]_T$ over a T time interval. The Kronecker product is used to combine the measurement matrices of two sensor nodes S1 and S2, S1(Φ)$\otimes$S2(Φ). Note that the computational cost of measurement matrix, Kronecker product, paillier encryption, and decryption are denoted as C_m, C_k, P_e, and P_d, respectively. The computational cost of a single compressive sensor device is estimated as follows.

$$C(S_r) = C_m + C_k + Pe \ldots\ldots\ldots\ldots..(11)$$

Similarly, the overall computational cost of r number of sensor devices $C(S_r)_{Total}$ are computed using equation (12).

$$C(S_r)_{Total} = r * C(S_r) \ldots\ldots\ldots.(12)$$

Further, the Sr nodes forward data to the corresponding CH node for aggregation and data transmission. The computation complexity of a single CH node is P_e. The total computational cost of intermediate CH nodes ($C(CH)_R$) involving in the routing path is estimated using the following equation (13).

$$C(CH_R) = (h - 1) * Pe \ldots\ldots\ldots\ldots(13)$$

Finally, the data is reached at BS, and the computational complexity of BS is Pd. Overall computational cost to deliver a single packet from an aggregator to BS, $C^2{}_{CH\to BS}$ is calculated in equation (14).

$$C^2{}_{CH\to BS} = \big((h - 1) * Pe\big) + p_d \ldots\ldots\ldots\ldots(14)$$

The PC2SR is a lightweight security system in which the computational cost of the key refreshing process is negligible. The overall computation complexity of measuring and transmitting the data is estimated as $\mathrm{C} = \mathrm{C}(\mathrm{S_r})_{\mathrm{Total}} + \mathrm{C^2}_{\mathrm{CH \rightarrow BS}}$.

6. CSHEAD Protocol Design

In this scheme, the neighbors monitoring mechanism for attack detection has been introduced. As the initial status, a trusted node is known as such by its neighbors in the network. If a node have been identified as an attacker by multiple neighbors, then the detected attacker is reported across the network to the sink so it should get isolated from the network. CSHEAD attack detection model is designed to send data only once to all nodes within same cluster including the cluster head or parent node and all nodes keep monitoring the activity of their neighbors. The secure aggregation and attack detection works as the following phases:

Phase 1: Every node start communicating to identify its neighbors. The Network architecture setup is completed in this phase. Formation of the Cluster: this is where the network structure is completed and the routing path is defined from nodes to the sink. Nodes desiring to become cluster heads send a Hello message to all neighbors which can accept to join the cluster. Each node will either respond to one of the requests to join a cluster or be the one sending the request to become the Cluster Head. At the end of this phase, the network will be splitted into clusters. This process will be repeated to enable every node to become a cluster head and thus better manage the network energy and life. This phase is similar to what happen in CSDA scheme as an initiation phase.

Phase 2: The sink generate public and private keys. The Sink has enough energy and is the responsible of generating the keys as decryption will happen in the Sink. Also the Sink is the most protected component of the network with no energy constraints. Then, distribution of public key from the sink to the nodes. The distribution will use the Cluster heads for routing the keys from sink to nodes.

Phase 3: Data aggregation in one cluster at round t. At this phase, all sensor nodes send their readings to their Cluster Head and to all nodes in the same cluster. At each round t, all nodes from same cluster receive encrypted data $Enc(M_i x_i(\mathrm{t}))$ that is being sent to the Cluster Head.

CH share its own encrypted data with its cluster members. Encrypted Data is now shared at cluster level and every node has received encrypted data from all nodes including Cluster Head.

Phase 4: Cluster Head multiply the received encrypted data from cluster members with its own encrypted reading. Same operation is done by the cluster members and they should all get same encrypted value $\prod_i Enc(M_i.x_i)$. Every node has received encrypted data from all other nodes within same cluster which it will multiply it with its own encrypted reading. The result should be the same value for all nodes. This value will the one sent by Cluster Head to the Sink.

Phase 5: Cluster Head forward the aggregated encrypted data to the Sink. During this phase, cluster members keep monitoring the Cluster Head (CH) activity and they compare the data sent by the CH to the Sink with their calculated data. If any difference is noticed then CH is reported as the attacker. The attacker is then isolated from the network. As the Cluster Head is the responsible of routing the data, then it has a critical role in forwarding correct data. If CH is compromised it may change the data and send other values to the Sink hence it’s important to monitor what data is being sent by CH to the Sink. Nodes are comparing their calculated values with the one sent by the CH. In case of a different value, they will report it to the Sink. Then the sink can detect the compromised node or Cluster Head and isolate it from the network and start phase 1 without the attacker.

Phase 6: The sink receive aggregated data from all cluster heads and calculate the $\prod_i Enc(M_i.x_i) = Enc(\sum_i M_i.x_i)$. As the sink has the decryption key, it will get original aggregated value $\sum_i M_i.x_i$ where M_i is only known by each node S_i, so even a compromised node won’t be able to reproduce easily encrypted data for another node even if they have same reading x_i and same public encryption key. The sink is using the aggregated value $\sum_i M_i.x_i$ for the whole network as a reference. If big change is detected over an acceptable threshold, then it can ask for getting raw data and detect if there is a compromised node. This phase allow to detect a node sending non expected data as its own reading. If the attacker is detected, the Sink will communicate the message to the network to isolate the attacker.

7. Simulation platforms

With the rapid development of wireless technologies, especially infrastructure less wireless networks such as WSNs, the evaluation and validation process has become difficult to implement in real environment because of its complexity, cost and long time required for fine tuning. To overcome such limitations, researchers have broadly adopted multiple simulators platforms to conduct out the implementation and validation of their finding. The main purpose of simulators is providing virtual platform with the required components to the distributed applications. For WSNs simulation, the detailed level of simulation environments determines the literalness of the experiments. But, it does not mean that the simulation environments are evaluated regarding to their detailed levels. Generally, simulation consumes too much resource. Thus, the challenge is to offer a balance between detail and performance. The existing simulators provide different levels of detail by providing different realistic layers according OSI model. The user should be aware of different components required to simulate each application in order to take the right decision regarding the simulator to use.

Nowadays, there is multiple simulation environment, which are cost and time effective. In the following paragraphs, we describe different simulation environments such as Network Simulator 2 (NS2), OPNET, QualNet and OMNet++ [2] and discuss their pros and cons in simulating WSNs.

7.1 Network Simulator 2

The network simulator NS2 is an open source platform basically developed by U.C. Berkeley/LBNL and extended to support WSN features by UCB Daedalus, CMU Monarch projects and Sun Microsystems. It is a discrete even object-oriented software dedicated for networking research area. The architecture of NS2 is mainly based on OSI model, its core engine is written in C++ and Otcl and simulation scenarios are TCL based. In Figure 3.5, we present the basic architecture of NS2. NS2 provides users with an executable command (ns), which takes the name of TCL simulation as argument. At the end of simulation, a trace file is generated. The trace file is then used to plot graph and/or to create animation.

NS2 consists of two key languages: C++ that constitutes the backend of the simulation and Object-oriented Tool Command Language (OTCL) constitutes the frontend that sets up simulation by configuring the objects and scheduling discrete events. The backend and frontend are connected together using TCL.

NS2 provides a large number of built-in C++ classes. It is recommended to use the C++ classes to set up a simulation through a TCL simulation script. However, advance users may find these objects insufficient. They need to develop their own C++ classes and use an OTCL configuration interface to put together objects instantiated from this class [129].

NS2 allows users to implement different protocols. Besides newly created protocols, all the internal protocols can be modified according to the need of users. Simulation in NS2 is achieved in three steps:

- First step, the user implements its protocol by writing C++ and OTCL codes to the NS2 source base.
- Second step, the simulation is described using an OTCL script.
- Last step, the user runs its protocol in NS2 and collects results by using trace files, which are generated by NS2.

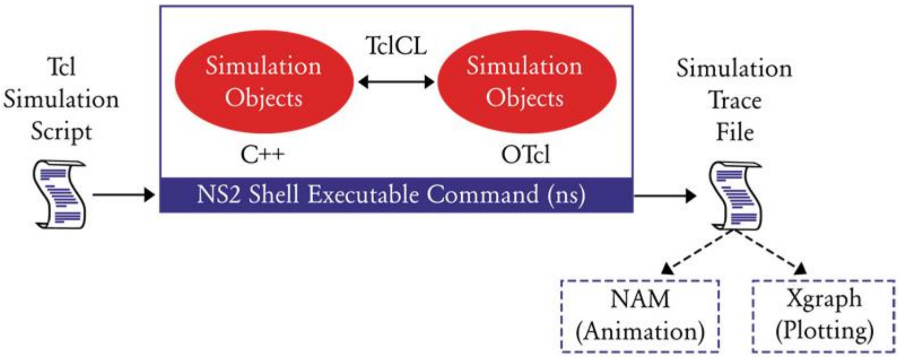


Figure 3.5 Basic architecture of NS2[31]

7.2 Network Simulator 3

NS3 is the successor of NS2 which last version 2.35 was released on 4th November 2011. NS3 is also open source and still getting enhancements from the community which contribute to the longevity of the project. It is suitable for research and education and last version ns-3.35 was released on October 1, 2021 as a result of contributions from nineteen authors[32]. As simulation will never be as good as real thing, simulators are constantly working on additional improvements and bug fixes. This allows to reproduce scenarios which are close to reality and even at large scale with cheap cost compared to real test beds.

The popularity of NS2 remains high as lot of documentation is available. Below, the differences between NS3 and NS2:

[31] Introduction to network simulator

[32] Network Simulator https://www.nsnam.org/releases/ns-3-35/

- NS3 is written using C++ while NS2 is a combination of C++ and the scripting language TCL thus, NS3 compilation time is not a concern.
- NS3 support python as scripting language which is not the case in NS2.
- NS3 improves memory management as it frees memory used to store the packets while NS2 never reuse or re allocate the memory until it get terminated.
- NS3 check the parameters required during execution and doesn't store unnecessary parameters unlike NS2 which can't prevent unnecessary parameters from being stored.
- NS3 includes other improvements making its total computation time lower when compared to NS2.
- NS3 is more difficult than NS-2 to implement and simulate a protocol or execute a scenario in WSN.

7.3 OPNet

The OPNet simulator is a commercial simulation platform, designed by OPNet technologies Inc. and sold to Riverbed Company. It offers modeling and simulation of networks devices, protocols and communication facilities. It supports the simulation of all type of wired networks and various wireless networks, such as IEEE 802.11. OPNET has the capability to implement new algorithms that are based on existing components [130].

7.4 QualNet

The QualNet simulator is a commercial tool, developed by Scalable Network Technologies Corporation. It uses a layered architecture model comparable to the TCP/IP model. According to that model, data moves between adjacent layers, communicate via application programming interface (APIs), from top to bottom, the Application, Transport, Network, Link (MAC) and Physical Layers. The mobility of nodes in QualNet is implemented via waypoint model, group mobility model, pedestrian mobility model, and file-base mobility model [131].

7.5 OMNet++

The OMNet++ simulator is an open-source, discrete event simulation environment. It based on C++ class library and offers networking modeling, multiprocessor and distributed systems simulation environment. The idea behind the development of OMNet++ was to offer a powerful simulation tool, which could be used freely by academics for education purposes and license based for commercial researches purposes [132][133].

7.6 Simulators selection criteria

Each simulator tool has its own advantages and disadvantages in different aspects. Various protocols can be implemented and simulated in these simulators. Although these simulators offer excellent simulation environments, almost all of them does not accurately reflect details related to real-life experiments.

Since real investigation of WSNs using experimentation platforms (test beds) is costly and not enough flexible to carry out all scenarios, Software-based simulation took place as an effective alternative and a widely used tool. Therefore, the choice of a simulator should be driven by the requirements of real application under study.

Table 4.1 point out some criteria that could help to select the simulator to use. For instance NS2 is the most popular network simulator in the network research community because it’s free for use and offer a high level of granularity. NS2 cannot easily handle large-scale simulation scenarios because it consumes very large amount of memory ones network size reach 500 nodes does. Simulators that support parallelism facilities could be useful in large-scale simulation scenarios such as OPNet. OPNet simulator offers various facilities in networking simulation but it still restricted for instance to commercial projects and some few academic projects. On the other hand OMNet++ could handle nearly 2000 nodes, it start taking place in academic network research projects.

Name	Popularity	License	Interface	Parallelism	Granularity
NS2	88.8%	Open source	C++/OTCL	No	Finest
OPNet	2.61%	Commercial	C	Yes	Fine
QualNet	2.49%	Commercial	Parsec (C-based)	SMP/Beowulf	Finer
OMNet++	1.04%	Free for academic use	C++	MPI/PVM	Medium

Table 3.1 Existing network simulators[33]

NS2 popularity remains very high even compared to NS3. This can be explained by the available documentation, protocols and the models within NS2 making it easy to use. NS2 protocols are free and easy to find publicly which major advantage for researches is when it comes to adding new protocols. In the following simulations, NS2 will be used as our choice was made to go for NS2 based on the previous assessment.

[33] Source: [131], An Overview of MANETs Simulation.

8. Performance Evaluation

The purpose of this section is to evaluate the performance of the proposed PC2SR and compare it with CSDA for secure data aggregation based on network simulator (NS2) platform [129], [134]. Simulation outputs will be used to assess the QoS level delivered by each secure data aggregation protocol under different density levels.

8.1 performance metrics

The evaluation of secure data aggregation performances is based on network throughput, average end to end packet delay, packet delivery ratio or packet loss ratio and routing overhead.

Aggregation performance will be assessed using number of nodes as input parameter. The number of nodes characterize how much density of nodes when the deployment area remain the same. Also attack detection accuracy will be assessed to determine the protocol performance and capability to detect adversary activities.

8.1.1 Throughput

Throughput (BW) is the measure of how fast we can actually send packets through network. The number of packets delivered to the receiver provides the throughput of the network. The throughput is defined as the total amount of data a receiver actually receives from the sender divided by the time it takes for receiver to get the last packet.

$$Bw = \frac{\sum R_p \times P_s}{\vartriangle t}$$

$$\vartriangle t = E_t - S_t$$

Where:

- R_p: received packets,
- P_s: packet size
- $\vartriangle$ t: transmission time
- E_t: End time
- S_t: Start time

8.1.2 Packet Delivery Ratio

Packet delivery ratio (PDR) is the ratio of the data packets delivered to the destinations to those generated by the CBR sources. It is the fraction of packets sent by the application that are received by the receivers.

$$PDR = \left(\frac{\sum D_p}{\sum S_p} \right) * 100$$

- D_p: packets delivered
- S_p: packets sent

8.1.3 Routing Overhead

Routing overhead (RO) is the number of routing packets transmitted per data packet delivered at the destination. Each hop-wise transmission of a routing packet is counted as one transmission. The routing overhead describes how many routing packets for route discovery and route maintenance need to be sent in order to propagate the data packets.

$$RO = \frac{\sum R_p}{\sum S_p}$$

- R_p: Routing packet

8.1.4 End-to-End Delay

End-to-End delay (D) indicates the total time taken by each packet to reach the destination. Average End-to-End delay of data packets includes all possible delays caused by buffering during route discovery, queuing delay at the interface, retransmission delays at the MAC propagation and transfer times.

$$D = \frac{\sum_{i=1}^{N} \vartriangle t[i]}{N}$$

- $\vartriangle$ t[i]: transmission time of packet [i]
- N: number of received packets

8.1.5 Detection Accuracy:

It is the percentage of the number of successfully detected attacks to the total number of attacks.

8.1.6 Energy Consumption:

It is the ratio of energy consumed by sensor devices to perform network operations.

8.2 Implementation

The trace file includes of about few hundreds to thousands of lines of information. We have used following AWK script to handle NS2 trace file output and calculate proposed new performance metrics.

8.2.1 NS2 Trace file format

The general format of each trace line is shown below:

0 ACTION: [s|r|D]: s -- sent, r -- received, D -- dropped

1 WHEN: the time when the action happened

2 WHERE: the node where the action happened

3 LAYER:

- AGT -- application,
- RTR -- routing,
- LL -- link layer (ARP is done here)
- IFQ -- outgoing packet queue (between link and mac layer)
- MAC -- mac,
- PHY -- physical

4 flags:

5 SEQNO: the sequence number of the packet

6 TYPE: the packet type

- cbr -- CBR data stream packet
- AODV -- AODV routing packet (control packet generated by routing)
- RTS -- RTS packet generated by MAC 802.11
- ARP -- link layer ARP packet

7 SIZE: the size of packet at current layer,

8 [a b c d]: mac layer details

- a -- the packet duration in mac layer header
- b -- the mac address of destination
- c -- the mac address of source
- d -- the mac type of the packet body

9 flags:

10 [……]: ip layer details
[source node ip: port_number
destination node ip (-1 means broadcast):
port_number
ip header ttl
ip of next hop (0 means node 0 or broadcast)]

8.2.2. Simulation parameters

The efficiency of the proposed PC2SR is analyzed using NS2. For performance analysis, the PC2SR is compared with the existing CSDA [135]. Simulation parameters are depicted in table 4.3.

A transmission range of 5.5 to 70 m can be attained for Sensor to sensor communications at 900 MHz when the sensor is on the ground. When the sensor is elevated 1 m above the ground, the transmission range can reach 140 m [136]. The simulation is based on a Transmission Range of 50 m which is a realistic assumption when the sensor are on the ground and without obstacles.

We used NS2.35 installed on Linux ubuntu 14.04.6 LTS to carry out our simulation. Moreover, TCL program has been used for simulation scenario and AWK script has been used to analyze trace files and calculate the quality of service metrics.

Parameters	Values
Simulation Area	100*100 m
Simulation Time	2000s
No. Of Nodes	80
Transmission Range	50 m
Queue Type	Droptail/Priority Queue
Queue Length	50 packets
Antenna Type	Omni Antenna
Propagation Type	Two ray ground
MAC layer protocol	802.15.4
Routing protocols	PC2SR
Transport agent	UDP
Application agent	CBR

Table 3.2 Simulation parameters

8.3. Simulation and Results discussion

8.3.1 Simulation review for CSHEAD and PC2SR

The simulation is repeated using 40, 50, 60, 70 ad 80 sensor nodes.

Step 1: Common for all protocols, as described in Figure 3.6, the network is created with 50 nodes. 0 is the Base Station (BS).

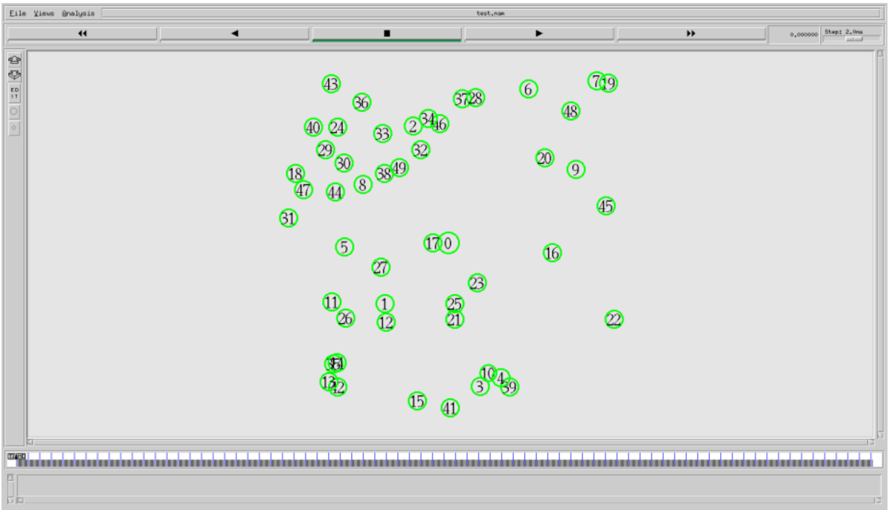


Figure 3.6 Network Architecture

Step 2: Cluster Heads (CH) are selected randomly using LEACH protocol.

Here node 13, 19, 32, 5, 27, 25, 14 are CH. This step is also the same for three schemes.

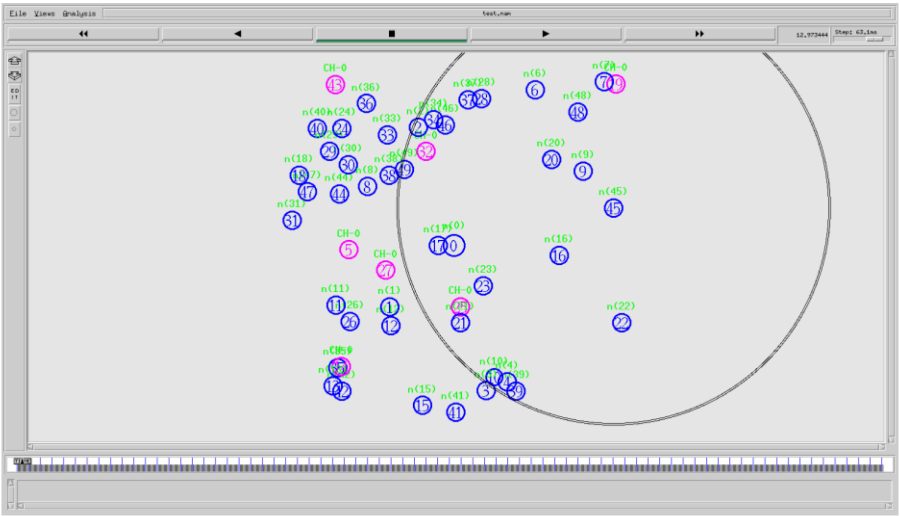
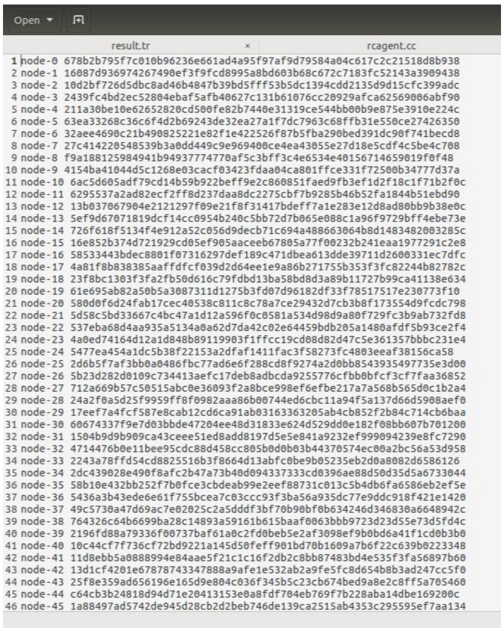


Figure 3.7 Cluster Head Selection

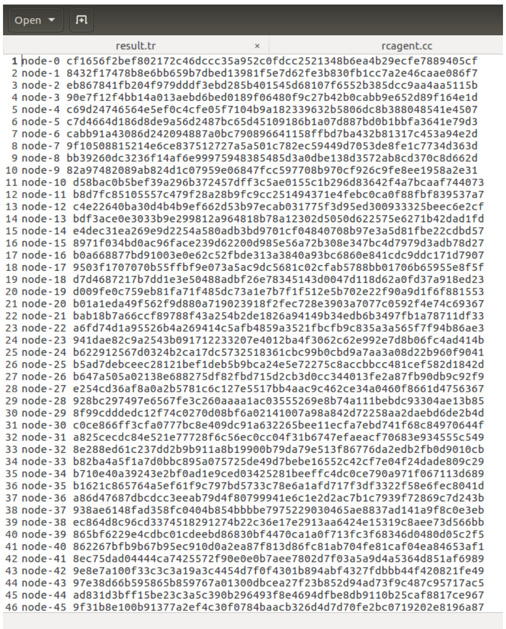
For encryption and decryption a semi-homomorphic encryption decryption method, Paillier cryptosystem is used. It is an asymmetric encryption decryption method, hence it will use public and private keys for data encryption and decryption respectively. In this the Paillier public and private keys are generated with the help of Paillier cryptosystem. Paillier public and private keys are only known to the respective nodes and BS. The public and private keys are updated based on rounds.

Step 3: In CSHEAD and PC2SR and for all the 50 nodes Paillier public and private keys are generated. The above figures shows the hex value of the Paillier public and private keys. Using these hex values, respective keys are generated.



```
node-0 678b2b795f7c010b96236e661ad4a95f97af9d79584a04c617c2c21518d8b938
node-1 16087d936974267490ef3f9fcd8995a8bd603b68c672c7183fc52143a3909438
node-2 10d2bf726d5dbc8ad46b4847b39bd5fff53b5dc1394cdd2135d9d15cfc399adc
node-3 2439fc4bd2ec52804ebaf5afb40627c131b61076cc20929afca62569006abf90
node-4 211a30be10e62652820cd500fe82b7440e31319ce544bb00b9e875e3910e224c
node-5 63ea33268c36c6f4d2b69243de32ea27a1f7dc7963c68ffb31e550ce27426350
node-6 32aee4690c21b490825221e82f1e422526f87b5fba290bed391dc90f741becd8
node-7 27c414220548539b3a0dd449c9e969400ce4ea43055e27d18e5cdf4c5be4c708
node-8 f9a188125984941b94937774770af5c3bff3c4e6534e40156714659019f0f48
node-9 4154ba41044d5c1268e03cacf03423fdaa04ca801ffce331f72500b34777d37a
node-10 6ac5d605adf79cd14b59b922beff9e2c860851faed9fb3ef1d2f18c1f71b2f0c
node-11 6295537a2ad82ecf2ff8d237daa8dc2275cbf7b9285b46b52fa1844b51ebd90
node-12 13b037067904e2121297f09e21f8f31417bdeff7a1e283e12d8ad80bb9b38e0c
node-13 5ef9d67071819dcf14cc0954b240c5bb72d7b065e088c1a96f9729bff4ebe73e
node-14 726f618f5134f4e912a52c056d9decb71c694a488663064b8d1483482003285c
node-15 16e852b374d721929cd05ef905aaceeb67805a77f00232b241eaa1977291c2e8
node-16 58533443bdec8801f07316297def189c471dbea613dde39711d2600331ec7dfc
node-17 4a81f8b838385aaffdfcf039d2d64ee1e9a86b271755b353f3fc82244b82782c
node-18 23f8bc1303f3fa2fb50d616c79fdbd13ba58bd8d3a89b11727b99ca41138e634
node-19 61e695ab82a50b5a3087311d1275b3fd07d96182df33f78517517e230773f10
node-20 580d0f6d24fab17cec40538c811c8c78a7ce29432d7cb3b8f173554d9fcdc798
node-21 5d58c5bd33667c4bc47a1d12a596f0c0581a534d98d9a80f729fc3b9ab732fd8
node-22 537eba68d4aa935a5134a0a62d7da42c02e64459bdb205a1480afdf5b93ce2f4
node-23 4a0ed74164d12a1d848b89119903f1ffcc19cd08d82d47c5e361357bbbc231e4
node-24 5477ea454a1dc5b38f22153a2dfaf1411fac3f58273fc4803eeaf38156ca58
node-25 2d6b5f7af3bb0a0486fbc77ad6e6f288cd8f9274a2d0bb8543935497735e3d00
node-26 5b23d282d0109c734413aefc17deb8adbcda9255776cfbb0bfcf3cf7faa36852
node-27 712a669b57c50515abc0e36093f2a8bce998ef6efbe217a7a568b565d0c1b2a4
node-28 24a2f0a5d25f9959ff8f0982aaa86b00744ed6cbc11a94f5a137d66d5908aef0
node-29 17eef7a4fcf587e8cab12cd6ca91ab03163363205ab4cb852f2b84c714cb6baa
node-30 60674337f9e7d03bbde47204ee48d31833e624d529dd0e182f08bb607b701200
node-31 1504b9d9b909ca43ceee51ed8add8197d5e5e841a9232ef999094239e8fc7290
node-32 4714476b0e11bee95cdc88d458cc805b0d0b03b44370574ec00a2bc56a53d958
node-33 2243a78ffd54cd8825516b3f8664d13abfc0be9b05235eb2d0a8082d6586126
node-34 2dc439028e490f8afc2b47a73b40d094337333cd0396ae88d50d35d5a6733044
node-35 58b10e432bb252f7b0fce3cbdeab99e2eef88731c013c5b4db6fa6586eb2ef5e
node-36 5436a3b43ede6e61f755bcea7c03ccc93f3ba56a935dc77e9ddc918f421e1420
node-37 49c5730a47d69ac7e02025c2a5dddf3bf70b90bf0b634246d346830a6648942c
node-38 764326c64b6699ba28c14893a59161b615baaf0063bbb9723d23d55e73d5fd4c
node-39 2196fd88a79336f00737baf61a0c2fd0beb5e2af3098ef9b0bd6a41f1cd0b3b0
node-40 10c44cf7f736cf72bd9221a145d50feff901bd70b1609a7b6f22c639b0223348
node-41 11d8ebb5a0888994e84aae5f21c1c16f2db2c8bb87483bd4e535f3fa56897b60
node-42 13d1cf4201e67878743347888a9afe1e532ab2a9fe5fc8d654b8b3ad247cc5f0
node-43 25f8e359ad656196e165d9e804c036f345b5c23cb674bed9a8e2c8ff5a705460
node-44 c64cb3b24818d94d71e20413153e0a8fdf704eb769f7b228aba14dbe169200c
node-45 1a88497ad5742de945d28cb2d2beb746de139ca2515ab4353c295595ef7aa134
```

Figure 3.8 Paillier private key



```
node-0 cf1656f2bef802172c46dccc35a952c0fdcc2521348b6ea4b29ecfe7889405cf
node-1 8432f17478b8e6bb659b7dbed13981f5e7d62fe3b830fb1cc7a2e46caae086f7
node-2 eb867841fb204f979dddf3ebd285b401545d68107f6552b385dcc9aa4aa5115b
node-3 90e7f12f4bb14a013aebd6bed0189f06480f9c27b42b0cabb9e652d89f164e1d
node-4 c69d24746564e5ef0c4cfe05f7104b9a182339632b5806dc8b388048541e4507
node-5 c7d4664d186d8de9a56d2487bc65d45109186b1a07d887bd0b1bbfa3641e79d3
node-6 cabb91a43086d242094887a0bc790896641158ffbd7ba432b81317c453a94e2d
node-7 9f10508815214e6ce837512727a5a501c782ec59449d7053de8fe1c7734d363d
node-8 bb39260dc3236f14af6e99975948385485d3a0dbe138d3572ab8cd370c8d662d
node-9 82a97482089ab824d1c07959e06847fcc597708b970cf926c9fe8ee1958a2e31
node-10 d58bac0b5bef39a296b372457dff3c5ae0155c1b296d83642f4a7bcaaf744073
node-11 b8d7fc85105557c479f28a28b9fc9cc251494371e4febc0ca0f88fbf839537a7
node-12 c4e22640ba30d4b4b9ef662d53b97ecab031775f3d95ed300933325beec6e2cf
node-13 bdf3ace0e3033b9e299812a964818b78a12302d5050d622575e6271b42dad1fd
node-14 e4dec31ea269e9d2254a580adb3bd9701cf04840708b97e3a5d81fbe22cdbd57
node-15 8971f034bd0ac96face239d62200d985e56a72b308e347bc4d7979d3adb78d27
node-16 b0a668877bd91003e0e62c52fbde313a3840a93bc6860e841cdc9ddc171d7907
node-17 9503f1707070b55ffbf9e073a5ac9dc5681c02cfab5788bb01706b65955e8f5f
node-18 d7d4687217b7dd1e3e50488adbf26e78345143d0047d118d62a0fd37a918ed23
node-19 d009fe0c759eb81fa71f485dc73a1e7b7f1f512e5b702e22f90a9d1f6f881553
node-20 b01a1eda49f562f9d880a719023918f2fec728e3903a7077c0592f4e74c69367
node-21 bab18b7a66ccf89788f43a254b2de1826a94149b34edb6b3497fb1a78711df33
node-22 a6fd74d1a95526b4a269414c5afb4859a3521fbcfb9c835a3a565f7f94b86ae3
node-23 941dae82c9a2543b091712233207e4012ba4f3062c62e992e7d8b06fc4ad414b
node-24 b622912567d0324b2ca17dc5732518361cbc99b0cbd9a7aa3a08d22b960f9041
node-25 b5ad7debceec28121bef1deb5b9bca24e5e72275c8accbbcc481cef582d1842d
node-26 b647a505a02138e688275df82fbd715d2cb3d0cc344013fe2a87fb90db9c92f9
node-27 e254cd36af8a0a2b5781c6c127e5517bb4aac9c462ce34a0460f8661d4756367
node-28 928bc297497e6567fe3c260aaaa1ac03555269e8b74a111bebdc93304ae13b85
node-29 8f99cdddedc12f74c0270d08bf6a02141007a98a842d72258aa2daebd6de2b4d
node-30 c0ce866ff3cfa0777bc8e409dc91a632265bee11ecfa7ebd741f68c84970644f
node-31 a825cecdc84e521e77728f6c56ec0cc04f31b6747efaeacf70683e934555c549
node-32 8e288ed61c237dd2b9b911a8b19900b79da79e513f86776da2edb2fb0d9010cb
node-33 b82ba4a5f1a7d0bbc895a075725de49d7bebe16552c42cf7e04f24dade809c29
node-34 b710e40a39243e2bf0ad1e9ced03425281beeffc4dc0ce790a971f067113d689
node-35 b1621c865764a5ef61f9c797bd5733c78e6a1afd717f3df3322f58e6fec8041d
node-36 a86d47687dbcdcc3eeab79d4f80799941e6c1e2d2ac7b1c7939f72869c7d243b
node-37 938ae6148fad358fc0404b854bbbbe7975229030465ae8837ad141a9f8c0e3eb
node-38 ec864d8c96cd3374518291274b22c36e17e2913aa6424e15319c8aee73d566bb
node-39 865bf6229e4cdbc01cdeebd86830bf4470ca1a0f713fc3f68346d0480d05c2f5
node-40 862267bfb9b67b95ec910d0a2ea87f813d86fc81ab704fe81caf04ea84653af1
node-41 8ec75dad04444ca7425572f90e0e0b7aee7802d7f03a5a9d4a5364d851af6989
node-42 9e8e7a100f33c3c3a19a3c4454d7f0f4301b894abf4327fdbbb44f420821fe49
node-43 97e38d66b595865b859767a01300dbcea27f23b852d94ad73f9c487c95717ac5
node-44 ad831d3bff15be23c3a5c390b296493f8e4694dfbe8db9110b25caf8817ce967
node-45 9f31b8e100b91377a2ef4c30f0784baacb326d4d7d70fe2bc0719202e8196a87
```

Figure 3.9 Paillier public key

Step 4: Compressive sensing:

Here, in PC2SR:

i. The sensed data is compressed to generate sparse matrix.

ii. Nodes encrypt data using paillier public key

iii. Cluster member nodes send data to their respective CH.

In CSHEAD:

i. The sensed data is compressed to generate sparse matrix.

ii. Nodes encrypt data using paillier public key.

iii. Cluster member nodes send data to their respective CH as well as other nodes within same cluster.

Example of Encryption using public keys.

*****PAILLIER ENCRYPTION (CM) *****

ENCRYPTED TO CIPHER TEXT:
2179103049351634685131944250575846374242175274670379178304104575222962158626993190826487928487836125264008633040813349992183829635822484434573434723010625

(CM) CIPHER TEXT CONVERTED TO BYTES ----> 0x5405be0

Step 5: Data Aggregation: The CH aggregate the data from the selected member nodes.

In PC2SR:

i. The received data is decrypted to generate original data.
ii. The decrypted data is added with some noise factor.
iii. Then this data and noise factor are encrypted and CH send it to BS.

In CSHEAD:

i. The received data is not decrypted.
ii. Data aggregation is applied on encrypted data.
iii. Cluster members do the same aggregation operation and should get the same value. This value will the one sent by Cluster Head to the Sink.
iv. CH send encrypted aggregated data to BS while other nodes listen and compare their calculated values with CH sent value. (In case of difference Cluster node notify BS : attack detection)

Step 7: Data Decryption: Here BS receive the data from CH and decrypted data.

Step 8:

In PC2SR:

BS will extract the data from the noise factor and take the sum of the noise factor to find out the false injection.

RECEIVED DATA AT BS FROM CH 24 DATA ----> 0x54153c0

BS DECRYPTED DATA ----> -11

Here received data are -11, 19. Then separate original data 11, 19 from the noise factor -1, 1. Sum of the noise factor is,

$$-1+1 \quad = \quad 0$$

Since sum of noise factor is zero the BS will process the data and consider the data are from genuine nodes.

In CSHEAD:

BS will decrypt and decompress data to get original readings. In case of other nodes notify attack, BS check notifications and isolate attacker which could be CH or sensor node.

8.3.2 Results and discussion

Simulation results presented in figure 3.10 demonstrate that PC2SR protocol; experience better detection accuracy than previous protocols for secure data aggregation.

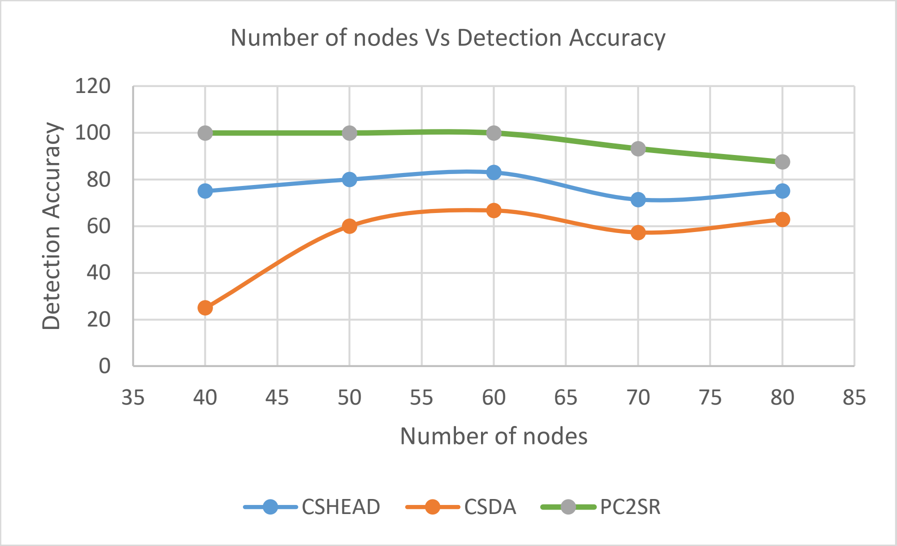


Figure 3.10 Number of Nodes Vs. Detection Accuracy

Figure 3.10 depicts the comparative results of detection accuracy of CSHEAD, PC2SR and CSDA obtained by varying the number of nodes from 40 from 80. The results of figure 3.10 show that CSHEAD offer better detection accuracy than CSDA but both have been outperformed by PC2SR. The new protocol PC2SR maintains the detection accuracy of 99.9% from 40 to 70, and it suddenly decreases the detection accuracy by 12.4% after the point 70. The reason behind this that the PC2SR appends a modified lightweight paillier cryptosystem with compressive sensing based data gathering mechanism to detect attack behaviors efficiently. However, some packets may lose due to congestion under a high node density scenario, and it diminishes the detection accuracy of PC2SR. Moreover, the proposed PC2SR attains enhanced secure routing performance than the existing CSDA. The proposed PC2SR utilizes zero noise factors with all transmitted data and strong fresh paillier keys for rich data security. For example, the detection accuracy of PC2SR is increased by 74.9% and 20% than the existing CSDA under the low and high-density scenario, respectively.

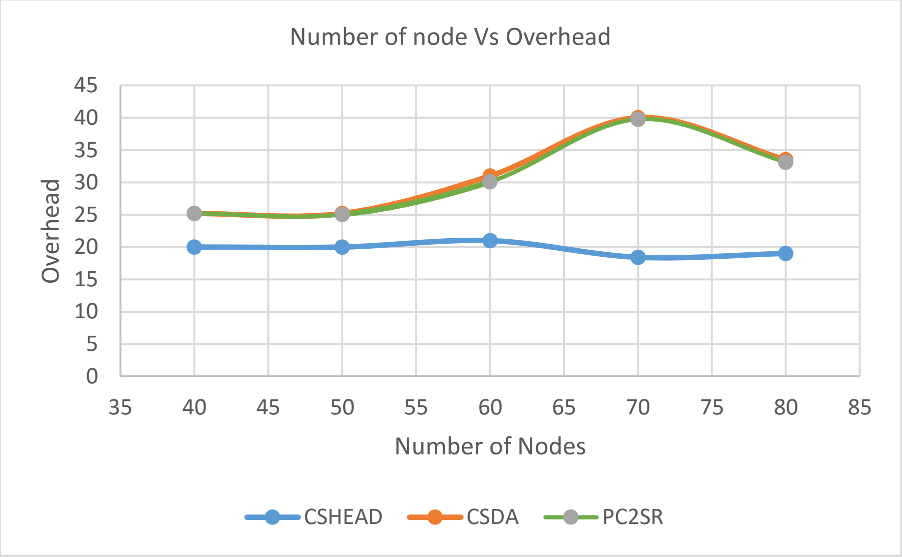


Figure 3.11 Number of Nodes Vs. Overhead

Overhead results of PC2SR and CSDA are plotted in figure 3.11. The results are obtained under different node density scenarios for comparatively analyzing the performance of PC2SR with CSDA and CSHEAD. Both CSDA and CSHEAD are having same overhead trend in different network densities. The compared mechanisms slightly escalate the overhead by varying the number of nodes from 40 to 80. The main reason is that both PC2SR and CSDA utilizes a significant amount of packets for key sharing and routing process. Thus, it increases the overhead by varying the number of nodes from low to high. For instance, the PC2SR accomplishes 0.1 and 0.48 of overhead for 40 and 80 node density scenarios. However, the overhead of PC2SR is less, when compared to existing CSDA. The slicing model of CSDA incurs some extra overhead in the network. Unlike CSDA, the PC2SR aggregates the measured data based on the compressive sensing model and delivers the BS with high security using lightweight mechanisms. In figure 3.11, the PC2SR reduces the overhead by 10.2% than the existing CSDA under a high node density scenario.

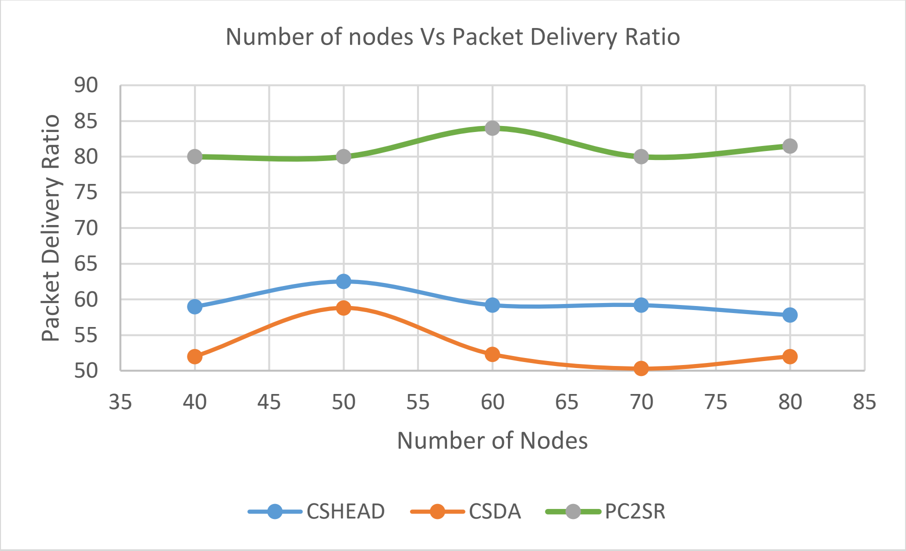


Figure 3.12 Number of Nodes Vs. Packet Delivery Ratio

Figure 3.12 illustrates the packet delivery ratio results of PC2SR, CSHEAD and CSDA estimated under diverse node density scenarios. Figure 3.12 clearly shows that the packet delivery ratio of PC2SR is more than that of the existing CSDA and its improved version CSHEAD. The proposed PC2SR constructs an inter-cluster tree for quick and efficient data delivery. Also, the utilization of strong security keys with light key refreshing mechanism prevents the attackers from packet dropping. For example, the PC2SR and CSDA attain PDR of 82% and 52%, respectively, when the numbers of nodes are 80 in the network.

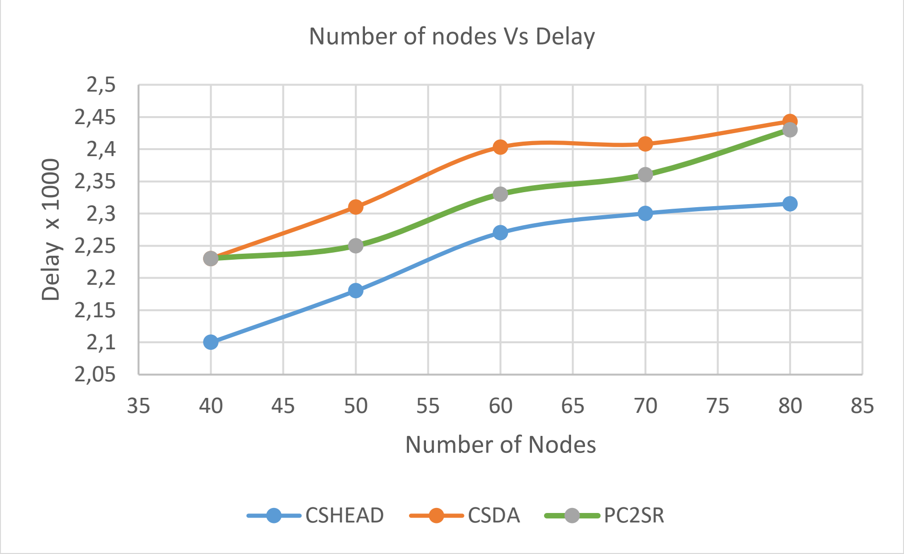


Figure 3.13 Number of Nodes Vs. Delay

Figure 3.13 portrays the comparative results of delay of PC2SR, CSHEAD and CSDA obtained by varying the numbers of nodes from 40 to 80. The 3 protocols improve the delay with varying the node density from low to high. This is because the number of nodes increases the routing complexities and thus incurs some delay in data packet delivery. For instance, the PC2SR attain a delay of 0.0023 and 0.0025 seconds of delay for 40 and 80 node densities, respectively. However, the delay of PC2SR is less than that of existing CSDA due to lightweight computations. CSHEAD offer better performance but all 3 scheme remain very close in term of delay. From figure 3.13, the PC2SR and CSDA attain a delay of 0.00245 and 0.00246 seconds, respectively, under a high node density scenario.

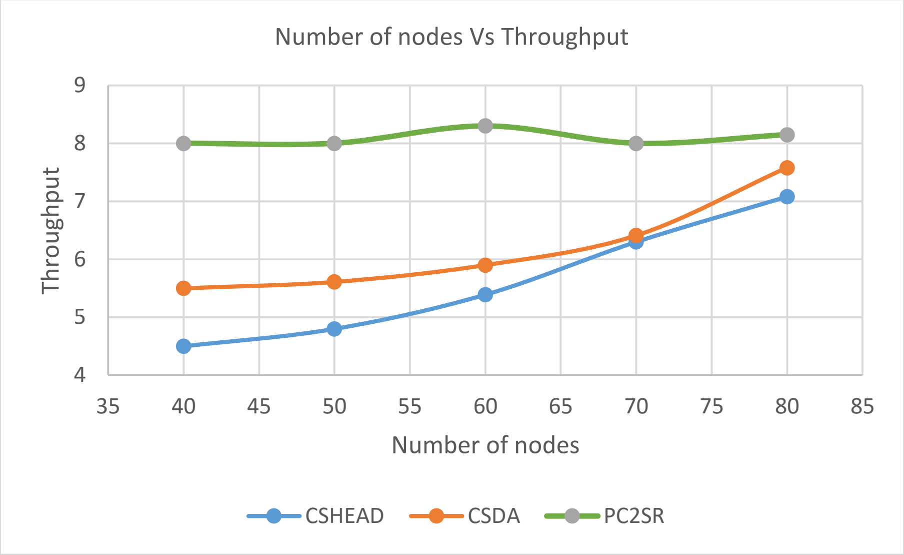


Figure 3.14 Number of nodes Vs. Throughput

Figure 3.14 shows the comparative results of throughput of PC2SR, CSHEAD and CSDA obtained by varying the numbers of nodes from 40 to 80. While PC2SR offers a fast and stable packets delivery through network, the number of packets delivered to the receiver by CSDA and CSHEAD is improving while increasing the network density.

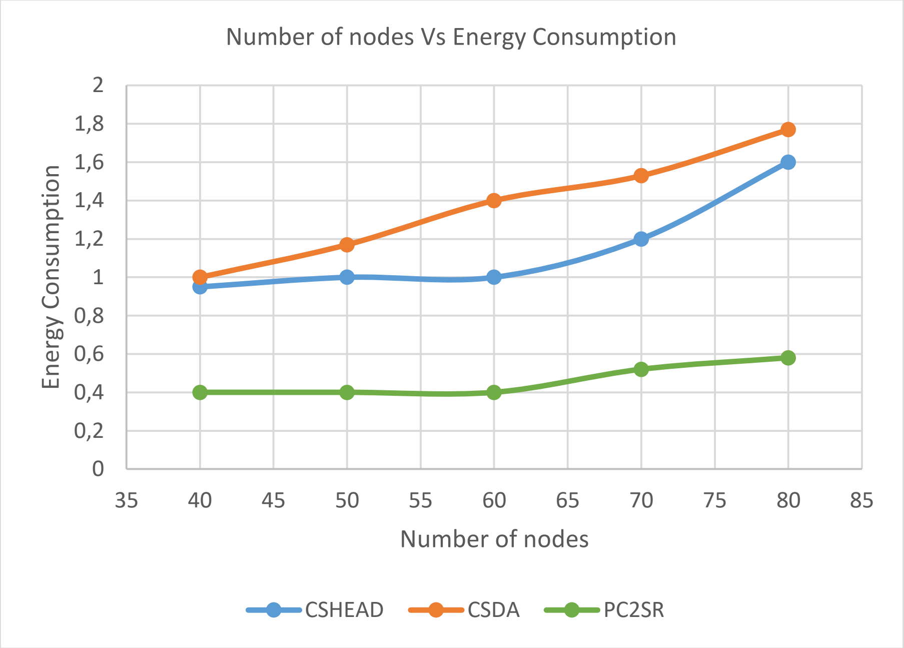


Figure 3.15 Number of Nodes Vs. Energy Consumption

The energy consumption results of PC2SR and CSDA are demonstrated in figure 3.15. To evaluate the efficiency of PC2SR, the results are obtained by varying the number of nodes from 40 to 80. The PC2SR maintains the energy consumption of nodes from 40 to 60 node density, and it increases the energy consumption after the point 50. The reason is that the number of nodes increases the energy consumption level, as the nodes have to communicate a large number of nodes to perform routing functions. Thus, it improves the node energy consumption level considerably. However, the PC2SR diminishes the node energy consumption level by integrating compressive sensing based data gathering model and lightweight security mechanisms. The Spatio-temporal sparse measurement matrix of PC2SR increases the data accuracy without consuming high energy at nodes. Moreover, the PC2SR reduces the energy consumption level of nodes by 51% than the existing CSDA.

9. Conclusion

In this chapter, we presented two novel secure data aggregation schemes which both provides better performance than the existing CSDA. The proposed schemes are using semi-Homomorphic cryptosystem for data encryption. This Cryptosystem allows arithmetic operations on encrypted data at the Cluster Heads without need to decrypt. The data confidentiality remain protected during all its route from the sensor node to the Sink. Sum function has been used for aggregation as it’s the case in many applications in practice. The first proposed CSHEAD scheme has been compared to the existing CSDA scheme. CSHEAD scheme reduce the number of communications in the network and has attack detection capability. Evaluation of both schemes is conducted under different network densities and measure the key metrics such as throughput, end-to-end delay, packet delivery ratio and routing overhead. CSHEAD improves all Network metrics and hence increases the network lifetime and overall performance. Also, Active Attack Detection is more accurate in CSHEAD scheme compared to CSDA.

The second evolution of CSHEAD is proposed via a lightweight, secure routing protocol named PC2SR. It detects various WSN security attacks and reduces the communication cost without compromising the data accuracy level. For that, the proposed PC2SR integrates a semi-homomorphic cryptosystem named as paillier cryptography with compressive sensing. Provision of strong paillier keys with key refreshing mechanism, the PC2SR significantly enhances the security level of nodes over WSN. The compressive sensing based data gathering within intra-cluster reduces the communication cost and also prolongs network lifetime without compromising the accuracy of data. Additionally, the zero noise factors with transmitted data increase the attack detection accuracy level at BS and also enhances the data accuracy level. Finally, the simulation results demonstrate the effectiveness of proposed PC2SR using performance metrics such as detection accuracy, overhead, packet delivery ratio, delay, and energy consumption.

Chapter 4

Anomalies Detection using Machine Learning Techniques in Wireless Sensor Networks

1. Introduction

The revolution of electronic components miniaturization process featuring wireless technologies influences our everyday life. With the popularity of smart phones, laptops and smart electronics in the post PC era, Information Technology devices have become affordable, more mobile, more distributed and more omnipresent in the society. It is now possible to construct an embedded system in the size of a wallet with an equivalent capacity of a PC from the 90s [1]. Such embedded systems can be supported by the extent down Windows or Linux operating systems. In this regard, the appearance of Wireless Sensor Networks (WSNs) is essentially the latest tendency of Moore's Law toward the miniaturization and ubiquity of computing devices.

Wireless sensor networks represent a special class of ad hoc networks [2]. They are made up of many smart sensor nodes of small sizes, limited power, at low-cost, and multifunctional (also called Nano computers). In principle, these network nodes have a spontaneous mode of organization because they are intended to be deployed quickly and arbitrarily in a space of interest. They are powered by a power unit (battery) of limited capacity. They can capture (or collect) physical quantities from the environment such as temperature, wind speed, relative humidity, etc. They are also able to detect real-world events, process data, and communicate with each other to bring the information collected to a collection point called sink node or Base Station (BS) [3, 4]. This information is then transmitted via a transport network to a processing center where possible analyzes, interpretations, and decision-making are carried out by an end-user.

At the present time, wireless sensor networks have become one of the hottest research areas due to their wide range of real-time applications like critical military surveillance, battlefields, building security monitoring, forest fire monitoring, and healthcare [5]. The design of these applications assumes that all the nodes involved are cooperative and trustworthy. However, this is not the case in real-world deployments, where nodes are exposed to different types of attacks and intrusions that can downright damage the proper functioning of the network and degrades system performance. Unfortunately, ensuring the security of this type of network against various malicious attacks activities is a difficult task, especially when the nodes are made up of inexpensive electronic devices with limited hardware capabilities [6].

Cryptographic algorithms require significant energy consumption, processing, and memory. In general, cryptographic and authentication algorithms provide the services of confidentiality, integrity, and authentication, although, with the use of only an encryption algorithm and security level management, it is difficult to guarantee that the data is legitimate and did not suffer any type of attack that extracted sensitive data, and use them for a malicious reason [7]. Whilst the cryptographic techniques solutions have been found to reduce cyber-attacks, but they have not eliminated them completely. Detection-based approaches are then proposed to protect WSNs from well-know and new cyber-attacks, as a second-line defense [8].

Intrusion detection systems (IDS) are one of the most flexible and useful tools to guard WSNs from known and unknown attacks. IDS observes and analyzes the events generated in the network to detect anything unusual and alert sensor nodes about the intruder [4, 6]. This concept was originally proposed by Anderson [9]. The strategies broadly utilized to develop IDS used for attack detection nowadays are vastly related to machine learning techniques. Most approaches, however, are based on online learning which requires all, or at least a sample, of historical data to be kept in memory. But, there are few approaches of detection anomalies that are actually learning models online.

The rest of this chapter is organized as follows: Section 2 is dedicated to problem statement and overview on attack detection in WSN. In section 3 we present previous work that tried to handle attack detection using machine learning techniques, and then we present the feature selection algorithms as long with incremental learning model with a cluster-based WSN. The simulation results used to evaluate the performance of the proposed model are presented in section 4. In section 6 is reserved to implementation details. The section 7 is devoted to highlight performance metrics used in the evaluation of proposed protocol. Finally, in section 8 we present simulation results and we discuss the performance of proposed model.

2. Attack detection in WSN

Solutions to security attacks against all networks, both wired and wireless involve three main components: Prevention, Detection, and Attenuation [29, 30]:

1. In the **preventive** step, the technique provides defense against the attack before it happens. It prevents the threat before it emerges.

2. The **detection** step kicks in when an attacker has found a way to bypass the preventive technique, which means that the nodes has been attacked. At this point, being aware of the attack that took place, the detection and identification of the nodes that are compromised is the aim of the detection phase.

3. The **mitigation** or attenuation step aims to suppress any attack detected during the detection phase by taking measures (revoking network routing tables) to secure the network from attackers and compromised nodes

Intrusion is any type of unwanted interruption activity in a network that is not allowed passively (e.g. information gathering, eavesdropping) or active (e.g. transmission of harmful packets, packet dropping, and hole attacks). When the prevention of such activity is not provided by the first line of defense of the WSN security, then the intrusion detection is carried out by network member nodes as a second line of defense to detect any suspicious behavior.

2.1 Intrusion detection system

Security professionals agree that an intrusion detection system (IDS) is a crucial part of the IT security infrastructure. IDS is a collection of resources, techniques and tools designed to detect any set of activities that attempt to compromise the integrity, confidentiality or availability of a resource. Confidentiality means ensuring that data is not disclosed to systems or unauthorized user or authorized users attempting to gain additional privileges or authorized persons who abuse their privileges. Integrity is the guarantee that data is preserved with regard to its meaning, completeness and intended use. Availability by ensuring that the data and the system are available whenever needed to be accessed by authorized users or systems [31].

After intrusion detection, it cannot act on it since this is a passive alert system, i.e. it only triggers alarms towards the controller. Intrusion detection systems provides some or all of the following information to the controller related to the intruder such as: node identification, location, the time and date of the intrusion, the intrusion activity (for example, active or passive) and the type of intrusion, etc. These information's are very useful for the third line of defense. (ie the process of responding to the threat through the mitigation step).

2.2 Intrusion detection approaches

The National Institute of Standards and Technology (NIST) [32] has classified intrusion detection into two main approaches: The misuse detection or also called approach by scenarios and Anomaly detection or also called behavioral approach. These two analysis methods are the important part of intrusion detection systems.

- **The knowledge-based approach**: In this approach, the records from previous attacks in the form of signatures and models are kept under form of database, these attack records are compared to the data received on the network (i.e. the system depends on prior knowledge of the attack signature). An intrusion is reported when the trace of a known attack is present in the database. However, this method cannot be used for the first attacks (for example, zero-day attacks) because it can only detect previously known threats that are present in the database. Hence, the signatures updates of new attacks and the knowledge update is a continuous requirement. On the other hand, the detection of abuse is very effective for detection of attacks with very low false positives (which is not an activity intruders), easy to develop and require less computer resources [33].
- **The behavioral approach**: In this approach, a model of normal network activity is first constructed. This model is called the normal behavior profile which will be used as a reference in the detection. During system monitoring, any significant deviation from the current behavior of the monitored network compared to normal behavior previously learned network gives rise to an attack [34]. The detection of anomalies do not require prior knowledge of the attacks (i.e. prior detection without knowledge). Therefore, it is able to detect new, unknown threats, which makes this method ideal for use against all attacks and infiltrations, including zero-day attacks. Anomaly detection allows the production of information that can be used to define signatures for a knowledge-based analysis [33]. However, the detection of anomalies is often

regarded as providing high false positive rate and requires learning phases to characterize normal behavior patterns, in addition to taking time to train the system.

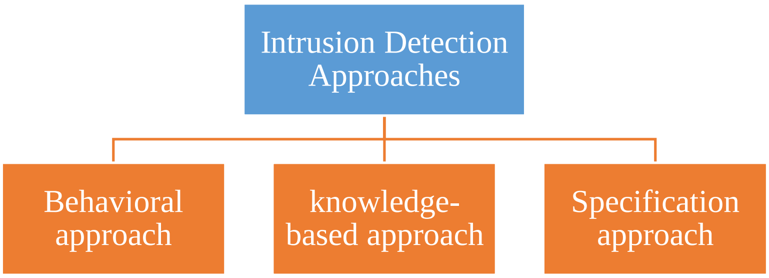


Figure 4.1 Intrusion Detection Approaches

The study of the two analytical approaches used by intrusion detection systems shows the existence of a completeness between these two methods. This completeness will overcome the drawbacks relating to each method of analysis. For this reason, it is preferable to adopt the two techniques in parallel in order to obtain an efficient intrusion detection system [35]. In this context, a third hybrid approach comes to exist by exploiting the advantages of the detection of abuses and the detection of anomalies.

- **The specification approach**: it's a hybrid between signature-based detection and anomalies detection where a set of specifications and constraints is developed to describe the normal behavior of the network [36]. Two detection mechanisms are usually combined, one to detect known attacks using signatures, the other to monitor traffic and detect deviations from normal network behavior as identified during learning phase. This methodology was introduced in [37], which offer the possibility of detecting unknown attacks, while presenting a low false positive alarm rate.

2.3 Anomalies detection techniques for WSN

An intrusion detection system based on the detection of anomalies monitors the network activities to classify them as normal or abnormal. The system proceed with the construction of behavior patterns for normal network activities and the observation of significant deviations of the current activity of the network compared to the established normal form [38]. In general, the detection of anomalies consists of two phases:

- **A learning phase**: The system learns the normal behavior of the network. It creates the normal network profile from the collected data.

A detection phase: The system compares the current audit trails or the network traffic with profiles to see if there is no intrusive activity. If the difference between the profile and the audit trails is significant, an alarm is triggered. In order to be able to formalize the normal behavior of a system, various approaches have been used. This section will be devoted to a general presentation of these approaches.

2.3.1 Statistical approach

In the methods of detecting statistical anomalies, the system creates profiles of behavior. As a general rule, two profiles are kept for each object: the current profiles and stored behavior profiles. The detection system updates the current profiles and regularly calculates an anomaly score by comparing the current profiles with the stored behavior profile. If the anomaly score is greater than a certain point, the intrusion detection system generates an alert [38].

2.3.2 Machine learning approach

The main purpose of using machine learning is the automatic extraction of characteristics related to normal activities that are critical for the detection of anomalies. Based on the audit data, the machine learning model tries to identify rules to define normal behaviors. These rules will be used to determine whether newly observed events are abnormal or not.

2.3.3 Data mining approach

The data mining approach is applied when there are many more normal observations than abnormal observations in the data. It consists in constructing a profile for normal behavior and also use the normal profile to detect anomalies which deviate significantly from the normal profile.

2.4 Anomaly detection based on machine learning

Human beings learn through life experience and machines follow the human instructions without change. Today machines can learn past data to help anticipate trends, future changes and opportunities. Machine Learning (ML) has been introduced as a subset discipline of artificial intelligence, characterized like a survey that allows personal computers to learn knowledge without being explicitly programmed as referenced by Arthur Samuel in 1959 stating that automatic learning focuses fundamentally on expectations [39].

Machine learning techniques have been introduced into a large number of applications to execute several tasks, including classifying, re-regressing, and estimating density in various fields of application, e.g. bioinformatics, speech recognition, spam detection, computer vision, fault detection and networks advertising [40].

2.4.1 Machine Learning and Anomalies detection modes

Depending to which extent labels are available, fault detection techniques can operate in one of the following three modes:

Supervised anomaly detection: The supervised anomaly detection approach uses predictive modeling for anomalies detection. The process involves the establishment of a normal model based on the specific data set. This mode defines both normal and abnormal data instances. Once the model is trained, it is applied to the new data set to uncover anomalies or to make predictions.

Semi-supervised anomaly detection: The semi-supervised anomaly detection approach involves the development of a combined model based normal and abnormal data. Typically, this combination will contain a very small amount of labeled data and a large amount of untagged data. This mode is more used because it is not an easy task to model all types of abnormal data.

Unsupervised anomaly detection: The unsupervised approach does not involve the development of a data model. The anomaly detection techniques based on this approach assume that there is no high risk of detecting anomalies in the data. Due to this approach, a high false alarm occurs.

2.4.2 Machine Learning and Online learning

Online learning (also known as incremental learning or out-of-core) [41] is a machine learning method that is used to construct a learnable model for an effective detection classification based on real time data coming from several sensors.

By avoiding traditional machine learning techniques (also called offline or batch learning) which require a lot of time and computation to process the data, the online model uses only the data provided previously and can therefore require frequent updates of the model manually on more recent data, then deploy the resulting model again whenever the system's normal behavior changes.

Online learning is efficient and adaptable for data. Online Learning is efficient in term of data because once the data is consumed, it is no longer necessary. Technically, this means that it is not needed to store the network data. Online learning is adaptable because it makes no assumptions about the distribution of systems. As the distribution of data changes or deviates, due for example to a change in the network behavior, the can adapt on the fly to keep pace with trends in real time.

3. Related work

WSNs are not excluded from the intrusion attacks and security threats, which lead to data privacy leaks or a decrease network its performance and efficiency. This motivates the growing research efforts to build efficient intrusion detectors for wireless sensor networks adapted to their specific characteristics. Various studies proposed machine learning solutions for intrusion detection systems to detect intrusion in WSNs. The existing intrusion detection methods mainly include online learning approaches such as Support vector machines, Random Forest, Artificial neural network, Decision tree and other methods. In literature there are only few works that aim to use online learning as an approach to benefit from the advantages of those techniques.

Almomani et al. [10] have developed a new specialized WSN dataset, and the collected dataset is called WSN-DS. It contained regular network traffic and several DoS (Flooding, Gray hole, Black hole and scheduling attacks) scenarios in WSN. It is created based on LEACH protocol, which is one of the most popular hierarchical routing protocol in WSNs, Using the network simulator NS2 to collect data. A (WEKA) data-mining toolbox was used for implanting artificial neural network (ANN) to detect the 4 attacks and classify them. The results were classified using both 10 folds cross-validation and holdout splitting methods. In their research work, the authors opted for a mechanism using the algorithm ANN trained by WSN-DS. This mechanism achieved a high classification of DoS attacks excluding the Gray hole attack since the detection rate is very low compared to the others.

Dong et al. [11] proposed an intrusion detection model based on information gain ratio and Bagging algorithm for detecting DoS attacks in a cluster-based WSNs. The authors used information gain ratio to reduce unnecessary features. The Bagging algorithm was used to construct an ensemble algorithm to train a set of C4.5 decisions trees in the aim of improving them. The proposed model was implemented by using both NSL-KDD and WSN-DS dataset separately to examine the performance of the model. This method provides enhanced performance than other methods.

Abdullah et al. [12] have studied a set of machine learning techniques for detecting DoS attacks with an Intrusion Detection System (IDS) applied for WSNs. Support vector machine (SVM), Naïve Bayesian, Random Forest, and Decision Tree (j48) classifiers were implemented with the WEKA data mining tool using WSN-DS as a dataset. From the results of this study, the SVM classifier has the upper hand in detecting intrusions with a high detection rate compared to the other techniques.

Sindhu et al. [13] have constructed a new lightweight IDS aimed for detecting anomalies in WSNs based on DT classification algorithm in WSN. For the implementation the authors chose to use Kddcup'99 as a dataset for relevant data. The model is based on three steps, in the first step feature selection method was implemented to remove irrelevant features for better results. The important features then were used in a wrapper based feature selection algorithm to identify suitable subset. The last step was adapting the learning paradigm neurotree in IDS. The authors claim that applying the right features with neurotree is a promising strategy for intrusion detection. Indeed, the model presented higher detection accuracy.

Pachauri et al. [14] have examined machine learning techniques, classification and regressions algorithms on real medical dataset with their proposed framework to detect faults and anomalies. The framework combines random forests algorithm for classification jobs and additive regression techniques for prediction jobs for anomaly detection in medical WSNs. The authors declare that their approach gives more accurate results than other existing fault detection mechanisms and both these algorithms perform much better than other previous research techniques.

Cauteruccio et al. [15] proposed a novel approach for Short-long term anomaly detection in heterogeneous wireless sensor networks based on machine learning and multi-parameterized edit distance, their method is performed by applying the analysis of edge and cloud on real data, which has been developed inside residential building and then deformed with a set of fake impairments. The obtained results show that the proposed method can self-adapt to the environment variations and correctly identify the anomalies.

Bosman et al. [16] have proposed a new lightweight framework for online anomaly detection in Internet of Things (IoT) applications including WSNs based on ensembles of incremental learners. Their decentralized approach was able to perform better than each individual centralized online learner alternatives to detect anomalies, even in an environments with little a priori knowledge, ending that ensemble schemes are feasible for practical implementation. The implementation used various large synthetic and real-world datasets, the evaluation of the proposed model was based on the prediction accuracy and confusion matrix metrics.

Bosman et al. [17] have also proposed a decentralized anomaly detection system for WSNs using unsupervised online learning approach. Central techniques have several drawbacks and for that an implementation with a range of real-world network deployments and data-sets were used to detect anomalies, also reducing energy and spectrum consumption by incorporating neighborhood information approach and using Recursive Least Squares (RLS) to learn linear models.

Rassam et al. [18] have introduces a variation of PCA called the One-Class Principal Component Classifier (OCPCC) for local and unsupervised anomaly detection, taking into consideration the energy consumption in WSN that uses the Candid Covariance free Incremental Principal Component Analysis (CCIPCA) algorithm to detect the intrusions as they occur. The implementation used GSB as a dataset and the approach is divided into two phases, the online phase a PCA model is trained using normal data collected from each sensor to build the normal behavior model, the online detection phase when the sensor nodes classify every packet as either normal or abnormal according to the threshold specified in a global normal model (GNM). The normal PCA model is updated and retrained with new mean and standard deviation of the new data. The proposed model achieved 96% as Detection accuracy with 7.2% as False Detection Rate.

Martins et al. [19] have proposed an online anomaly detection using a Least Squares-Support Vector Machine algorithm (LSSVM), under the form of a Reproducing Kernel Hilbert Space (RKHS) with Radial Basis Function (RBF) kernel, along with a sliding window-based learning technique. The proposed model was tested on a dataset generated with a virtual system that was used to assess the performance of the proposed approach. Simulation results have shown the out-performance of the proposed approach.

Myint et al. [20] have proposed one classifier known as Incremental Learning Algorithm (ISVMM), which is based on a support vector machine with Mahalanobis distance [12]. In this, a prediction is done by using SVM and is going to reduce steps required for calculation and complexity of the algorithm. This is achieved via ending a support set, error set, remaining set and providing a hard and soft decision. Time is then saved for repeatedly training the dataset. The authors used for simulation KDD Cup99 as a dataset to check the performance of the system. The proposed ISVMM model can predict well on all of the 41 features without reducing the dimensionality of the dataset.

In conclusion, authors have proposed different approaches to enhance the performance of intrusion detection systems in wireless sensor networks. Offline approach is efficient in terms of performance but generate a huge amount of load to regularly update the model once deployed. On the other hand online learning classifiers have not been thoroughly addressed in the literature. Our aim is to provide an intrusion detection model compatible with the characteristics of WSN.

In this chapter, we will use in our experiment a specialized dataset WSN-DS in order to classify four types of DoS attacks: Black hole, Gray hole, Flooding, and Scheduling among normal network traffic. We will compare three feature selection methods with the online Passive-Aggressive classifier, and examine his performance with each method and end the pair that achieve better results. We will propose an intelligent, efficient, and learnable model using the online classifier Passive-Aggressive with applying feature reduction, and ensuring that the model is compatible with the characteristics of WSN.

4. Research Methodology

4.1 Feature selection

Feature selection in the WSN are data pre-processing methods that are intended to reduce the number of irrelevant input variables to those that are believed to be most related to the intrusion attack. Relevant features have decisive effects on the output of classification, such as increasing the efficiency of the IDS and reducing energy consumption.

Some predictive model problems have a large number of variables that can slow the development and training of the models and require a large amount of system memory. Furthermore, the performance of some models can degrade when including input variables that are not relevant to the target variable. Eliminating non relevant variables and keeping only more suitable features supports to:

- Reduce the dataset size
- Reduce the time needed to train the model
- Reduce the risk of over fitting
- Reduce the risk of data misleading
- Consume less resources which is more suitable for online learning phase

For this reason, we choose 3 feature selection methods that are known to give a great performance and obtain the effective features with our model.

4.1.1 Chi-squared

The Chi-squared statistic is used to compute a score between the target and the numerical variable and only select the variable with the maximum chi-squared values. In feature selection, chi-squared measures the independence of features with respect to the class. The initial belief is that the feature and the class are independent before computing a score [25]. A score with high value means the existence of a high-dependent relationship.

$$X^2 = \frac{(\text{Observed frequency} - \text{Expected frequency})^2}{\text{Expected frequency}} \qquad (1)$$

Where:

Observed frequency = Number of observations in class.

Expected frequency = Number of expected observations of class if there was no relationship between the feature and target attribute.

4.1.2 Information gain

Information gain used in determining relevant features from a set of features. Before the start of the learning process it is used to rank and select the top features to reduce the feature size based on information theory. The greater the information gain, the more important the features are. Prior to ranking, the entropy value of the distribution is measured to determine the uncertainty of each feature according to their relevance in determining different class [26].

The entropy of the variable X is defined as follows:

$$H(X) = \sum_{i=0}^{n} P(x_i) log_2(P(x_i)) \qquad (2)$$

Where $P(x_i)$ is the value of prior probabilities of X . The entropy of X after observing values of another variable Y is defined as:

$$H(X/Y) = \sum_j P(y_j) \sum_i P(x_i/y_j)\, log_2 P(x_i/y_j) \qquad (3)$$

Where in Equation.3 $P(y_j)$ is the prior probability of the value y_j of Y, and $P(x_i/y_j)$ is the posterior probability of X given Y. The information gain is defined as the amount by which the entropy of X decreases to reflect additional information about X provided by Y and is defined as:

$$IG(X/Y) = H(X) - H(X/Y) \qquad (4)$$

According to this measure, if $IG(X/Y) > IG(Z/Y)$ then the feature X and Y are more correlated than feature Y and Z. The ranking used to select the most important features is calculated using the equation 4.

4.1.3 Information gain ratio

The purpose of information gain ratio is to improve the bias of information gain towards features with large diversity value [27].

$$GR(X) = \ IG(X)/H(X) \tag{5}$$

Where $GR(X)$ is the information gain ratio of feature X. Gain ratio takes number and size of branches into account when choosing an attribute and corrects the information gain by taking the intrinsic information of a split into account. Intrinsic information is the information about the class is disregarded.

4.1.4 Other Feature Selection Algorithms

Recursive Feature Elimination (RFE): This technique works in rounds and in each round it detects and removes features with high correlation. RFE is suitable for a dataset with high co-linearity and dependencies between its features. RFE can identify and eliminate these features. As first step, RFE identify important features and then rank them based on their importance. Second step is to eliminate the weakest features. RFE class mainly takes two arguments, the classifier and the number of features to be selected. Logistic Regression classifier can be used in several frameworks as it takes comparatively less time for training the model. RFE train the model using the classifier provided and calculate the accuracy by eliminating the unwanted features. RFE takes more time to compare to Univariate Feature Selection because it trains the model until the end of the loop.

Linear Discriminant Analysis (LDA): LDA is a widely used feature selection technique. This technique mainly removes the redundant and dependent features from the dataset. Mainly, it consists of three main stages. In the first stage, it calculates the difference between the averages of different classes. This difference is known as Between Class Variance. In the second stage, it calculates the difference between the average and the sample values of each class. This difference is known as Within Class Variance. In the third stage, it selects the features that have greater Between Class Variance and less Within Class Variance. LDA is also broadly used in the field of bio-informatics and chemistry. Consider that the probability density function of x with mean vector μi and variance-covariance matrix (same for all populations) is multivariate normal in population πi. For this scenario, the normal function of probability density is calculated as given below:

$$\mathrm{P(X / \pi i)} = \frac{1}{(2\Pi)^{p/2}|\sum|1/2} \exp\left[-\frac{1}{2}(\mathrm{X} - \mu\mathrm{i})'(\mathrm{X} - \mu\mathrm{i})\right] \tag{6}$$

Principle Component Analysis (PCA): PCA is widely used in unsupervised learning. The approach of this technique is very simple, but it can make a fair difference between the accuracy of the model trained when applied to all the features. Initially, it calculates the covariance of data points and arranged them in a matrix form. Further, it calculates the Eigen Vector and Eigen Value of that matrix. Then, it arranges all Eigen Vectors according to their Eigen Values in descending order. Furthermore, it selects the most promising features for training the model. It converts the original dataset into the selected number of Eigen Vectors. PCA is also used in the field of medical science and chemistry. It is also used to reduce the distortion from a graph. The below formula describes the covariance computation of X & Y. Where X and Y are matrices of m, and P is a linear transformation. X is the original data set point, and Y represents the Re-deployment dataset.

$$\mathrm{PX} = \mathrm{Y} \tag{7}$$

$$\mathrm{cov(X, Y)} = \frac{1}{(\mathrm{n}-1)} \sum_{i=0}^{n} (\mathrm{Xi} - \bar{\mathrm{x}})(\mathrm{Yi} - \bar{y}) \tag{8}$$

4.2 Online Machine Learning

Online learning (also known as incremental or out-of-core learning) [28] is a method of machine learning that builds a learnable model for effective classification in the real-time detection where the data is coming from multiple sensors. Avoiding the traditional machine learning techniques (also known as online or batch learning) that require time and great computing to process the data, the model use only previously provided data. The online learning may require frequent model updates manually on newer data and then deploying the resulting model again every time the normal system behavior change.

Incremental learning techniques allowing the model to be updated after receiving new data and making the model learns over time, without requiring a large historic data-set to be kept in memory. The structure of incremental learning classifiers can detect novel intrusions and handle concept drift in a dynamic network that is changed over time.

4.3 Online Passive-Aggressive Algorithm

Online Passive-Aggressive Algorithm (PA) is a family of online learning algorithms (for both classification and regression) proposed by Crammer at al. [29] it is similar to support vector machine classifier and can be considered as the online version of it. The idea is very simple and its performance has been proved to be superior to many other alternative methods like Online Perceptron and Margin-infused relaxed algorithm (MIRA). The PA classifier learns from streaming data and try to find hyper-planes to separate the instance into halves.

Algorithm 2 Passive-aggressive classifier

```
1: **Initialize:**
    $w_t$ ← (0, …, 0)
2: **for** t = 1, 2...., T **do**
    Receive instance: $x_t \in R^n$
    Predict: $y_t \leftarrow sign(w_t . x_t)$
    Observe correct label: $y_t \in \{-1, +1\}$
    Suffer loss: $l_t \leftarrow max\{0, 1 - y(w_t . x_t)\}$
    Set: $\tau \leftarrow l_t / ||x_t||^2$
    Update: $w_{t+1} \leftarrow w_t + \tau_t . y_t . x_t$
3: **end for**
```

Where w_t the weight of the vector on round t is, y_t is the signed margin. $w_t + 1$ is set to be the projection of w_t in the half-space of vectors that achieve a hinge-loss of zero. The algorithm is passive when a correct classification occurs having a hinge-loss as zero otherwise the classifier adjusts its weight vector for each misclassified training sample it receives. The Passive Aggressive classifier tries to get correct classification and updates the classification model. Online passive-aggressive (PA) classifier gives better results as it's learning rate does not decrease with respect to time making him a suitable solution for WSN.

4.4 Proposed approach

To improve the efficiency of security in WSN we propose a novel smart attack detection approach, which aim to reduce the side effect of new attacks. This is designed to detect existing known attacks as well as new attacks or Day-Zero attacks using a combination of offline and online learning.

Our proposal is based on information gain ratio and online passive aggressive algorithm, which we name as (ID-GOPA). Instead of using the classical method of offline learning only which provide good attack detection for known ones but will fail easily to detect new attacks. WSN Intrusion detection model based on information gain ratio and online passive aggressive algorithm, the shortened form is ID-GOPA. The main purpose of the proposed model, Figure 4.2, is to apply the study of the online classifier for the streaming data of the network. ID-GOPA inspects all events circulating in the network by observing abnormal activities and it consists of two phases: the offline and the online phase.

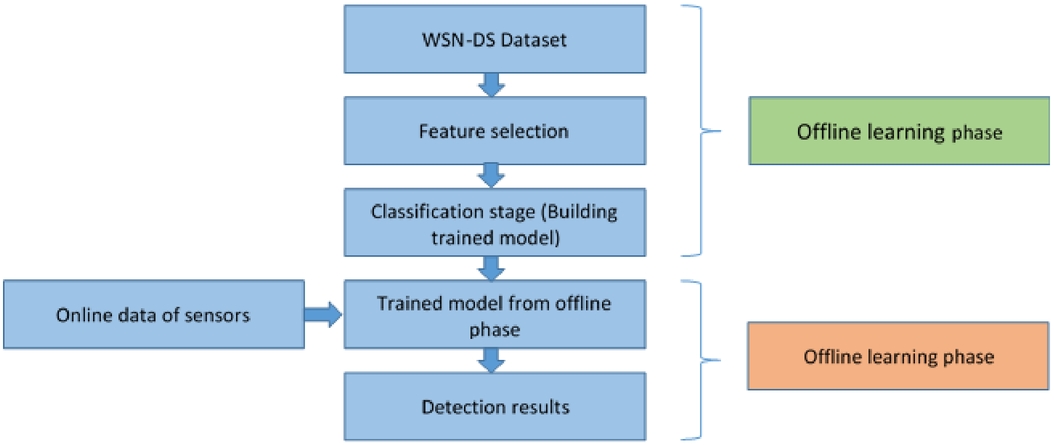


Figure 4.2 The proposed structure of the ID-GOPA model.

- In the offline phase (training dataset), the model is trained by the online classifier Passive Aggressive to be more familiar and more learnable for existing activities in the network flow, where the processed and labeled learning records are introduced to build a learnable model capable of being tested.
- In the online phase, using the trained model from the offline phase with the same preprocessing engine selecting only the relevant attribute based on information gain ratio algorithm and classifying every packet as either normal or attack in real-time detection.

5 Experiments and analysis

5.1 WSN-DS Dataset

The experiment uses a simulated wireless sensor network-detection system (WSN-DS) dataset developed by Almomani et al. [10], and the network simulator NS-2 was used to simulate wireless sensor network environment based on the LEACH routing protocol to collect data from network and preprocess it to generate 23 features identifying the state of each sensor. Then simulates four different types of Denial of Service (DoS) attacks: Black hole, Gray hole, Flooding, and Scheduling. The simulation parameters are summarized in Table 4.1. Only 19 features including the class label were in the dataset file as showed in Table 4.2. This dataset was generated as an IDS dataset to apply machine learning techniques in order to detect and classify DoS attacks. The data distribution is shown in Figure 4.2.

The technical characteristics of the computer adopted in the implementation phase are:

- Central Processing Unit: Intel(R) Core(TM) i7-4610M CPU @ 3.00GHz 3.00 GHz
- Random Access Memory: 8 GB
- Operating System: Windows 7 Pro 64-bit

Parameters	Values
Simulation Area	100*100 m
Simulation Time	3600s
No. Of Nodes	100
Routing protocol	LEACH
Number of clusters	5
Sink Location	(50,175)
Size of packet header	25 bytes
Size of data packet	500 bytes
Transmission Range	50 m
Queue Type	Droptail/Priority Queue
Queue Length	50 packets
Antenna Type	Omni Antenna
Propagation Type	Two ray ground
MAC layer protocol	802.15.4
Transport agent	UDP
Application agent	CBR

Table 4.1 WSN Simulation parameters.

Feature number	Symbol	Feature name	Description
1	id	Node Id	A unique ID number of the sensor node
2	Time	Time	The run-time of the node in the simulation
3	Is CH	Is CH	Describes if the node is a CH or not
4	Who CH	Who CH	Cluster head ID
5	Dist To CH	Distance to CH	Distance between node and CH
6	ADV S	ADV CH sends	Number of the advertise CH’s broadcast messages sent to nodes
7	ADV R	ADV CH receives	Number of advertise messages received by the nodes from CH
8	JOIN S	Join request send	Number of join request messages sent by the nodes to the CH
9	JOIN R	Join request receive	Number of join request messages received by CH from nodes
10	SCH S	ADV SCH sends	messages of TDMA schedule broadcast sent to the nodes
11	SCH R	ADV SCH receives	Number of scheduled messages received by the CH
12	Rank	Rank	Node order in TDMA scheduling

13	DATA S	Data sent	Number of data packets sent from the node to its CH
14	DATA R	Data received	Number of data packets received by the node from the CH
15	Data Sent To BS	Data sent to BS	Number of data packets that are sent from node to the BS
16	dist CH To BS	Distance CH to BS	Distance between CH and BS
17	send code	Send code	The sending code of the cluster
18	Consumed Energy	Energy consumption	Energy consumed
19	Attack type	Attack type	Type of attacks or normal traffic

Table 4.2 Features of the WSN-DS Dataset

The Attack Type	Training Set(60%)	Testing Set(40%)
Normal	204174	135892
Grayhole	8653	5943
Blackhole	5999	4050
Scheduling	4007	2631
Flooding	1963	1349
Sum	224796	149865

Table 4.3 The dataset separated into training and testing set

The dataset was split into training and testing sets with 60% of the data were used as the training data set and 40% of the data were used as the test data set. The number of observations in these two sets is presented in Table 4.3.

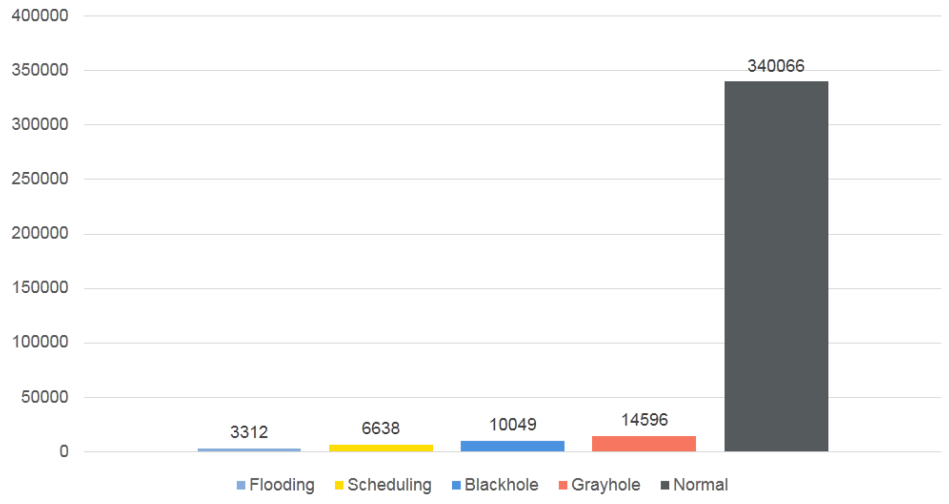


Figure 4.3 Distribution of WSN-DS dataset

5.2 Performance evaluation

The results of this study are evaluated according to four criteria, namely accuracy (ACC), precision (PR), f1-score (F), and recall (RE). All these criteria's take a value between

0 and 1. When it approaches 1, the performance increases, while when it approaches 0, it decreases. These performance evaluation metrics are computed as:

Accuracy (Acc): It estimates the ratio of the correctly recognized records to the entire test dataset. Accuracy serves as a good measure for the test dataset that contains balanced classes and defined as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \tag{9}$$

Precision (PR): Or Positive Predictive value (PPV) presents the ratio of the correctly classified data as the attack to all data classified as the attack and it is defined as follows :

$$\text{Precision} = \frac{TP}{TP + FP} \tag{10}$$

Recall (RE): It is also called as True Positive Rate (TPR) or (Sensitivity), it estimates the ratio of data classified as an attack to all attack data:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{11}$$

F1-Score (F): or F1-Measure represents the harmonic mean of the two matrices Precision and Recall. This concept is used to express the overall success:

$$\text{F1} - \text{Score} = \frac{2\ X\ PR\ X\ RE}{PR\ +RE} \tag{12}$$

The four values summarized below are used:

- True positive (TP) represents the number of correctly classified attack instances (correct detection).
- True negative (TN) represents the number of successfully classified normal data as being normal (correct rejection).
- False positive (FP) is the number of wrongly classified attack instances.
- False negative (FN) is the number of wrongly classified normal instances.

5.3 Feature selection

In this study, feature selection methods chis-squared, information gain and information gain ratio are used with WSN-DS data for comparative analysis of each one using the online

PA algorithm to verify the results of the selection. The selected features and performance are shown in Table 4.4.

Feature selection method	Feature selection result	Irrelevant feature	Accuracy (%)
Information gain ratio	3,5,6,7,8,9,10,11,12,13,15,16,17,18	1,2,4,14	95,69
Chi-squared	1,2,3,4,5,6,7,9,10,12,13,14,15,16	8,11,17,18	90,68
Information gain	1,3,4,5,6,7,8,10,11,12,13,15,17,18	2,9,14,16	90,68

Table 4.4 Comparison of feature selection methods

Through the experiment using Accuracy (Acc) as an evaluation index, it can be seen from Table 4 that the algorithm chi-squared and Information gain both of their Accuracy reached 90.68%. In the other hand when using the information gain ratio the Accuracy has gotten higher to 95.69% which means that the proposed WSN intrusion detection model has a better classification accuracy when choosing the information gain ratio as the attribute selection method. The selected feature set is $S = \{3,5,6,7,8,9,10,11,12,13,15,16,17,18\}$.

5.4 Simulation results and Discussion

Table 4.6 presents the detection performance of the proposed WSN intrusion detection model for normal scenario and with the attacks, using WSN-DS dataset, such as Black hole, Gray hole, Flooding, and Scheduling.

Normal	134353	1473	19	0
Gray hole	1212	664	4067	0
Black hole	0	2168	1882	0
Scheduling	357	5	2	2267
Flooding	433	63	1	0

Table 4.5 Online Passive Aggressive confusion matrix

The whole accuracy amount is 96%, as we analyze each class label to observe each individual performance, we see that the detection performance of normal cases is very high compared to abnormal cases with detection rate of 99%. And about the detection rate of Scheduling, Gray hole and Flooding attacks is 86%, 68% and 63%, respectively, on the other hand the Black hole attack got the worst detection rate with a percentage of 46%.

We have to take in consideration that the Passive-Aggressive classifier learns from streaming data, as the classifier learns, he tries to train the model better and also the accuracy gets better over time/iterations. A general trend is that individual online learning classifiers are characterized by a reduced recall and from the results obtained high accuracy was achieved in the task of classifying four DoS attacks to determine whether the protocol in its normal mode or exposed to any type of attack.

	PR	RE	F
Normal	0,99	0,99	0,99
Gray hole	0,52	0,68	0,59
Black hole	0,73	0,46	0,57
Scheduling	1,00	0,86	0,93
Flooding	0,93	0,63	0,75
Weighted avg.	0,96	0,96	0,96
Overall Accuracy		0,96	

Table 4.6 Online Passive Aggressive Results

5.5 Comparison with existing offline learning algorithms

For the purpose of comparison, four algorithms of machine learning are considered, namely SVM, NB, RF, DT to compare our work with using the same dataset (WSN-DS) and the specific results of performance comparison were measured and compared using Recall, precision, F1-score and accuracy index.

	PR	RE	F	Acc (%)
SVM	0,88	0,92	0,90	89%
Naïve Bayes	0,94	0,85	0,88	94%
Random Forest	0,94	0,85	0,88	94%
Decision Tree	0,94	0,94	0,93	94%
ID-GOPA	0,96	0,96	0,96	96%

Table 4.7 Performance comparison of various methods of WSN intrusion detection models

As can be seen from Table 7, the RE of the proposed method reaches 96%, which is higher than that of SVM, NB, DT, and RF. Among them, the Accuracy of SVM method is 89%, which is the smallest compared with the above methods. The mathematical reason behind the low accuracy is that SVM works better for small dataset and in our experimentation, we have used large dataset. Again, we have mentioned that the accuracy of Online PA classifier is high because it works better for large stream dataset and we have used large dataset for our experimentation in addition it gives better results, as it's learning rate does not decrease with respect to time since most of the online algorithms their concepts might change through time.

6 Conclusion

Providing security services in WSN based on intrusion detection systems to identify attacks with high accuracy is a challenging task. In this chapter we have presented an intelligent intrusion detection model based on incremental machine learning. The model determines the presence of an intrusion, and classifies the type of attack in real-time based on a cluster WSN network topology. The proposed model ID-GOPA efficiently detects intrusion, and avoids the resource waste. It uses information gain ratio as a feature selection to reduce the number of parameters and processing load. Feature selection is an important factor which improves the performance of the model with the passive-aggressive algorithm as an incremental learning machine. The simulation results shows an overall accuracy of 96% which means our model is very accurate compared to offline models. Our model is applicable to any application which makes it advantageous compared to existing models which are specific to their applications.

As future improvement work, we plan to go further by combining an ensemble of algorithms to detect anomalies. This would theoretically result in a better detection accuracy since there are more algorithms working together to overcome each other's limitations.

Conclusion & Perspectives

The main aim of this thesis was the design and the implementation of novel approach for secure data aggregation and attack detection in WSNs to protect data confidentiality and overcome the impact of resource scarcity on the overall network performances. The key research questions were centered on managing the negative effect of encryption on computation and attack detection scheme on network performance and the most effective strategy that can be adopted to improve data security and network performances under harsh and vulnerable deployment environment for WSN.

The priceless applications of WSN, especially its use for saving human life in disaster area and health monitoring as well as building monitoring where generally traditional communication systems might be out of service, were the fundamental motivation behind addressing the issue of security in WSN environment.

Along this thesis, we examined working process of multiple existing security schemes in the literature. We investigated the data aggregation attacks and protection as one critical step in forwarding sensor readings to the base station. We also, generalized the attack detection model by the usage of machine learning. The outcomes of this study showed up that Homomorphic encryption once combined with compressive sensing can allow designing novel secure data aggregation schemes that outperform existing ones. Also adopting a combined approach between offline and online learning lead to getting a novel protocol with better results in term of attacks detection. In addition, we simulated three secure data aggregation routing protocols, CSDA, CSHEAD and PC2SR using NS2.

The main contribution in this thesis is the design and implementation of two novel secure data aggregation schemes called CSHEAD and PC2SR, which are an improvement of CSDA scheme. The proposed CSHEAD and PC2SR schemes uses homomorphic encryption and compressive sensing techniques in its working process to provide confidentiality and attack detection with better network performance than the existing CSDA. In CSHEAD data is not decrypted at cluster head level but all nodes are getting the same data and contribute to monitor

Cluster Head communication to detect attacks. In PC2SR the performance is better than CSHEAD even if cluster head decrypt data and add a noise parameter before encrypting and sending to the base station. This means the decryption is consuming less energy overall than keeping all nodes at receiving mode which is the case in CSHEAD.

Performance evaluation was carried out based on six performance metrics such as delay, throughput, routing overhead, PDR, energy consumption and attack detection accuracy and has shown that CSHEAD and PC2SR outperforms CSDA.

The second contribution, is about designing a new intrusion detection system based on machine learning named ID-GOPA. The new system is combining the advantages of offline and online learning approaches and proceed with feature selection process to reduce the data and computation load. This intrusion detection system is not related to any specific attack and hence can be trained to detect know attacks as well as continue incremental learning online to detect new attacks. The proposed model ID-GOPA efficiently detects intrusion, and avoids the resource waste. It used information gain ratio as a feature selection model with the passive-aggressive algorithm to reduce the parameters and processing load. System evaluation was measured and compared using four metric which are recall, precision, f1-score and accuracy index. The new ID-GOPA model outperform existing attack detection systems such as SVM, NB, DT, and RF.

Likewise every system, the proposed approach has some limitations. The PC2SR secure data aggregation scheme is designed based on LEACH protocol. Moreover, performance evaluation has been carried out using network simulator 2. The main limitation of this platform resides in buffering size, which is not scalable for high density network and large simulation time. In our case, simulation with 1000 nodes and 2000s of simulation time has caused simulator freezing. Consequently, large scale network with long time simulation still unfeasible with NS2.

As perspectives of the present work, we identified three main ideas, which need to be explored. First, we plan to go further by combining an ensemble of machine earning algorithms to detect anomalies. This would theoretically result in a better detection accuracy since there are more algorithms working together to overcome each other's limitations.

Second, develop and use a test bed for performance evaluation purposes in order to assess CSHEAD and PC2SR behavior under realistic conditions.

Finally, extend performance simulation using other type of traffic flows such real time applications, streaming, TCP based applications.

References

[1] T. G. Robertazzi, *Introduction to Computer Networking*. Cham: Springer International Publishing, 2017. doi: 10.1007/978-3-319-53103-8.

[2] J. Loo, J. Lloret Mauri, and J. H. Ortiz, Eds., *Mobile ad hoc networks: current status and future trends*. Boca Raton, Fla.: CRC Press, 2016.

[3] P. Sharma., "Evolution of Mobile Wireless Communication Networks-1G to 5G as well as Future Prospective of Next Generation Communication Network," *Int. J. Comput. Sci. Mob. Comput.*, vol. 2, no. 8, pp. 47–53, Aug. 2013.

[4] S. K. Sarkar, T. G. Basavaraju, and C. Puttamadappa, *Ad hoc mobile wireless networks: principles, protocols, and applications*, Second edition. Boca Raton: CRC Press, 2013.

[5] S. Ahmadi, *Mobile WiMAX: A Systems Approach to Understanding IEEE 802.16m Radio Access Technology*. Academic Press, 2010.

[6] G. R. M. Reddy, M, and Kiran, *Mobile ad hoc networks: bio-inspired quality of service aware routing protocols*. 2017. Accessed: Jul. 03, 2018. [Online]. Available: http://dx.doi.org/10.1201/9781315368641

[7] R. I. Jenkins, *Bluetooth and wireless local area networks: security guides*. 2013. Accessed: Jun. 14, 2018. [Online]. Available: http://public.eblib.com/choice/publicfullrecord.aspx?p=4773185

[8] T. Qiu, N. Chen, K. Li, D. Qiao, and Z. Fu, "Heterogeneous ad hoc networks: Architectures, advances and challenges," *Ad Hoc Netw.*, vol. 55, pp. 143–152, Feb. 2017, doi: 10.1016/j.adhoc.2016.11.001.

[9] S. Zeadally, R. Hunt, Y.-S. Chen, A. Irwin, and A. Hassan, "Vehicular ad hoc networks (VANETS): status, results, and challenges," *Telecommun. Syst.*, vol. 50, no. 4, pp. 217–241, Aug. 2012, doi: 10.1007/s11235-010-9400-5.

[10] K. Singh and A. K. Verma, "Flying Adhoc Networks Concept and Challenges," *Encycl. Inf. Sci. Technol. Fourth Ed.*, pp. 6106–6113, 2018, doi: 10.4018/978-1-5225-2255-3.ch530.

[11] J. Yang, S. Liu, Q. Liu, and G. Qiao, "UMDR: Multi-Path Routing Protocol for Underwater Ad Hoc Networks with Directional Antenna," *J. Phys. Conf. Ser.*, vol. 960, no. 1, p. 012010, 2018, doi: 10.1088/1742-6596/960/1/012010.

[12] B. A. Sulaiman and M. M. Said, "Intelligent mobile ad hoc network management system," Thesis, Brunel University London, 2016. Accessed: Feb. 27, 2018. [Online]. Available: http://bura.brunel.ac.uk/handle/2438/14147

[13] IEEE Computer Society, LAN/MAN Standards Committee, Institute of Electrical and Electronics Engineers, and IEEE-SA Standards Board, *IEEE standard for Information technology-- telecommunications and information exchange between systems-- local and metropolitan area networks-- specific requirements: Part 11 : Wireless LAN medium access control (MAC) and physical layer (PHY) specifications : Amendment 6: Wireless access in vehicular environments*. New York: Institute of Electrical and Electronics Engineers, 2010. Accessed: Jul. 04, 2018. [Online]. Available: http://ieeexplore.ieee.org/servlet/opac?punumber=5514473

[14] J. Sanguesa *et al.*, “Sensing Traffic Density Combining V2V and V2I Wireless Communications,” *Sensors*, vol. 15, no. 12, pp. 31794–31810, Dec. 2015, doi: 10.3390/s151229889.

[15] İ. Bekmezci, O. K. Sahingoz, and Ş. Temel, “Flying Ad-Hoc Networks (FANETs): A survey,” *Ad Hoc Netw.*, vol. 11, no. 3, pp. 1254–1270, May 2013, doi: 10.1016/j.adhoc.2012.12.004.

[16] K. Kumari, B. Sah, and S. Maakar, “A survey: different mobility model for FANET,” *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 5, no. 6, 2015.

[17] S. Khan, A.-S. K. Pathan, and N. A. Alrajeh, *Wireless Sensor Networks Current Status and Future Trends*. Boca Raton: CRC Press, 2016.

[18] F. Liu and Y. Bai, “An overview of topology control mechanisms in multi-radio multi-channel wireless mesh networks,” *EURASIP J. Wirel. Commun. Netw.*, vol. 2012, no. 1, p. 324, Dec. 2012, doi: 10.1186/1687-1499-2012-324.

[19] M. Al-Shalabi, M. Anbar, T.-C. Wan, and A. Khasawneh, “Variants of the low-energy adaptive clustering hierarchy protocol: Survey, issues and challenges,” *Electronics*, vol. 7, no. 8, p. 136, 2018.

[20] D. Kandris, C. Nakas, D. Vomvas, and G. Koulouras, “Applications of Wireless Sensor Networks: An Up-to-Date Survey,” *Appl. Syst. Innov.*, vol. 3, no. 1, p. 14, Feb. 2020, doi: 10.3390/asi3010014.

[21] A. Tripathi, H. P. Gupta, T. Dutta, R. Mishra, K. K. Shukla, and S. Jit, “Coverage and connectivity in WSNs: A survey, research issues and challenges,” *IEEE Access*, vol. 6, pp. 26971–26992, 2018.

[22] T. Azzabi, H. Farhat, and N. Sahli, “A survey on wireless sensor networks security issues and military specificities,” in *2017 International Conference on Advanced Systems and Electric Technologies (IC_ASET)*, 2017, pp. 66–72.

[23] B. Prabhu, M. Pradeep, and E. Gajendran, “Applications of Wireless Sensor Networks in Battlefield Surveillance,” *Int. J. Technol. Res. Eng.*, vol. 4, no. 5, 2017.

[24] B. Prabhu, M. Pradeep, and E. Gajendran, “Enhanced battlefield surveillance methodology using wireless sensor network,” *Enhanc. Battlef. Surveill. Methodol. Using Wirel. Sens. Netw. January 25 2017 Multidiscip. J. Sci. Res. Educ.*, vol. 3, no. 1, 2017.

[25] N. Dey, A. S. Ashour, F. Shi, S. J. Fong, and R. S. Sherratt, “Developing residential wireless sensor networks for ECG healthcare monitoring,” *IEEE Trans. Consum. Electron.*, vol. 63, no. 4, pp. 442–449, 2017.

[26] R. Kashyap, “Applications of wireless sensor networks in healthcare,” in *IoT and WSN Applications for Modern Agricultural Advancements: Emerging Research and Opportunities*, IGI Global, 2020, pp. 8–40.

[27] M. Pule, A. Yahya, and J. Chuma, “Wireless sensor networks: A survey on monitoring water quality,” *J. Appl. Res. Technol.*, vol. 15, no. 6, pp. 562–570, 2017.

[28] “EEFFL: energy efficient data forwarding for forest fire detection using localization technique in wireless sensor network | SpringerLink.” https://link.springer.com/article/10.1007/s11276-020-02393-1 (accessed May 30, 2021).

[29] M. Dener, Y. Özkök, and C. Bostancıoğlu, “Fire Detection Systems in Wireless Sensor Networks,” *Procedia - Soc. Behav. Sci.*, vol. 195, pp. 1846–1850, Jul. 2015, doi: 10.1016/j.sbspro.2015.06.408.

[30] N. Varela, D.-M. Jorge L, A. Ospino, and N. A. Lizardo Zelaya, “Wireless sensor network for forest fire detection,” *Procedia Comput. Sci.*, vol. 175, pp. 435–440, Jan. 2020, doi: 10.1016/j.procs.2020.07.061.

[31] K. K. Khedo, Y. Bissessur, and D. S. Goolaub, “An inland Wireless Sensor Network system for monitoring seismic activity,” *Future Gener. Comput. Syst.*, vol. 105, pp. 520–532, Apr. 2020, doi: 10.1016/j.future.2019.12.025.

[32] “Modern Agriculture Using Wireless Sensor Network (WSN) | IEEE Conference Publication | IEEE Xplore.” https://ieeexplore.ieee.org/document/8728284 (accessed May 30, 2021).

[33] P. Ghosh, A. Gasparri, J. Jin, and B. Krishnamachari, “Robotic Wireless Sensor Networks,” in *Mission-Oriented Sensor Networks and Systems: Art and Science: Volume 2: Advances*, H. M. Ammari, Ed. Cham: Springer International Publishing, 2019, pp. 545–595. doi: 10.1007/978-3-319-92384-0_16.

[34] “Global Industrial Wireless Sensor Networks Market Trends, Growth Analysis 2028,” *MarketResearch.Biz*. https://marketresearch.biz/report/industrial-wireless-sensor-networks-market/ (accessed May 30, 2021).

[35] S. Gauhar Fatima, S. Kausar Fatima, S. Mohd Ali, N. Ahmed Khan, and S. Adil, “Methodologies and Challenges of WSN for IoT,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3534644, 2019. Accessed: May 30, 2021. [Online]. Available: https://papers.ssrn.com/abstract=3534644

[36] R. K. Dwivedi, M. Saran, and R. Kumar, “A survey on security over sensor-cloud,” in *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2019, pp. 31–37.

[37] C. Nakas, D. Kandris, and G. Visvardis, “Energy Efficient Routing in Wireless Sensor Networks: A Comprehensive Survey,” *Algorithms*, vol. 13, no. 3, Art. no. 3, Mar. 2020, doi: 10.3390/a13030072.

[38] T. Liu, J. Peng, J. Yang, G. Chen, and W. Xu, “Avoidance of energy hole problem based on feedback mechanism for heterogeneous sensor networks,” *Int. J. Distrib. Sens. Netw.*, vol. 13, no. 6, p. 1550147717713625, 2017.

[39] A. Lipare, D. R. Edla, and V. Kuppili, “Energy efficient load balancing approach for avoiding energy hole problem in WSN using Grey Wolf Optimizer with novel fitness function,” *Appl. Soft Comput.*, vol. 84, p. 105706, 2019.

[40] I. Tomić and J. A. McCann, “A survey of potential security issues in existing wireless sensor network protocols,” *IEEE Internet Things J.*, vol. 4, no. 6, pp. 1910–1923, 2017.

[41] M. Pawar and J. Agarwal, “A literature survey on security issues of WSN and different types of attacks in network,” *Indian J Comput Sci Eng*, vol. 8, no. 2, pp. 80–83, 2017.

[42] A. Lipare, D. R. Edla, and R. Dharavath, “Energy Efficient Routing Structure to Avoid Energy Hole Problem in Multi-Layer Network Model,” *Wirel. Pers. Commun.*, pp. 1–22, 2020.

[43] L. Tawalbeh, S. Hashish, and H. Tawalbeh, “Quality of Service requirements and Challenges in Generic WSN Infrastructures,” *Procedia Comput. Sci.*, vol. 109, pp. 1116–1121, Jan. 2017, doi: 10.1016/j.procs.2017.05.441.

[44] A. Ayyasamy and M. Archana, “Design and Analysis of QoS for Different Routing Protocol in Mobile Ad Hoc Networks,” in *Next-Generation Networks*, Springer, 2018, pp. 247–253.

[45] S. Balhara and P. Bhardwaj, “Bandwidth Constrained Multipath Routing Protocol for QoS Provision in MANETs,” *Trans. Netw. Commun.*, vol. 5, no. 3, Jun. 2017, doi: 10.14738/tnc.53.3282.

[46] G. K. Wadhwani and N. Mishra, “A Survey of Multicast Routing Protocols in MANET,” *IITM J. Manag. IT*, vol. 8, no. 1, pp. 42–50, 2017.

[47] N. Sharmin, A. Karmaker, W. L. Lambert, M. S. Alam, and M. S. T. Shawkat, “Minimizing the energy hole problem in wireless sensor networks: a wedge merging approach,” *Sensors*, vol. 20, no. 1, p. 277, 2020.

[48] L. Yadav and C. Sunitha, “Low energy adaptive clustering hierarchy in wireless sensor network (LEACH),” *Int. J. Comput. Sci. Inf. Technol.*, vol. 5, no. 3, pp. 4661–4664, 2014.

[49] T. Jamal and S. A. Butt, “Low-Energy Adaptive Clustering Hierarchy (LEACH) Enhancement for Military Security Operations,” *Proc J. Basic Appl. Sci. Res. ISSN*, pp. 2090–4304, 2017.

[50] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, “Energy-efficient communication protocol for wireless microsensor networks,” in *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, Jan. 2000, p. 10 pp. vol.2-. doi: 10.1109/HICSS.2000.926982.

[51] M. Al-Shalabi, M. Anbar, T.-C. Wan, and A. Khasawneh, “Variants of the low-energy adaptive clustering hierarchy protocol: Survey, issues and challenges,” *Electronics*, vol. 7, no. 8, p. 136, 2018.

[52] Y. Ma, Y. Guo, X. Tian, and M. Ghanem, “Distributed Clustering-Based Aggregation Algorithm for Spatial Correlated Sensor Networks,” *IEEE Sens. J.*, vol. 11, no. 3, pp. 641–648, Mar. 2011, doi: 10.1109/JSEN.2010.2056916.

[53] J.-H. Shin, J. Kim, K. Park, and D. Park, “Railroad: virtual infrastructure for data dissemination in wireless sensor networks,” in *Proceedings of the 2nd ACM international workshop on Performance evaluation of wireless ad hoc, sensor, and ubiquitous networks*, New York, NY, USA, Oct. 2005, pp. 168–174. doi: 10.1145/1089803.1089982.

[54] I. Solis and K. Obraczka, “Isolines: efficient spatio-temporal data aggregation in sensor networks,” *Wirel. Commun. Mob. Comput.*, vol. 9, no. 3, pp. 357–367, 2009, doi: 10.1002/wcm.551.

[55] M. Khademi Nori and S. Sharifian, “EDMARA2: a hierarchical routing protocol for EH-WSNs,” *Wirel. Netw.*, vol. 26, no. 6, pp. 4303–4317, Aug. 2020, doi: 10.1007/s11276-020-02328-w.

[56] O. Younis and S. Fahmy, “HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks,” *IEEE Trans. Mob. Comput.*, vol. 3, no. 4, pp. 366–379, Oct. 2004, doi: 10.1109/TMC.2004.41.

[57] H. Nyquist, “Certain Topics in Telegraph Transmission Theory,” *Trans. Am. Inst. Electr. Eng.*, vol. 47, no. 2, pp. 617–644, Apr. 1928, doi: 10.1109/T-AIEE.1928.5055024.

[58] C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948, doi: 10.1002/j.1538-7305.1948.tb01338.x.

[59] E. T. Whittaker, “XVIII.—On the Functions which are represented by the Expansions of the Interpolation-Theory,” *Proc. R. Soc. Edinb.*, vol. 35, pp. 181–194, ed 1915, doi: 10.1017/S0370164600017806.

[60] E. J. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006, doi: 10.1109/TIT.2005.862083.

[61] D. L. Donoho, “For most large underdetermined systems of equations, the minimal □1-norm near-solution approximates the sparsest near-solution,” *Commun. Pure Appl. Math.*, vol. 59, no. 7, pp. 907–934, Jul. 2006, doi: 10.1002/cpa.20131.

[62] “CS-PLM: Compressive Sensing Data Gathering Algorithm Based on Packet Loss Matching in Sensor Networks.” https://search.emarefa.net/en/detail/BIM-1216062-cs-

plm-compressive-sensing-data-gathering-algorithm-based-on (accessed Jun. 20, 2021).

[63] S.-H. Hsieh, T.-H. Hung, C.-S. Lu, Y.-C. Chen, and S.-C. Pei, "A Secure Compressive Sensing-Based Data Gathering System via Cloud Assistance," *IEEE Access*, vol. 6, pp. 31840–31853, 2018, doi: 10.1109/ACCESS.2018.2844184.

[64] C. Zhang, O. Li, X. Tong, K. Ke, and M. Li, "Spatiotemporal Data Gathering Based on Compressive Sensing in WSNs," *IEEE Wirel. Commun. Lett.*, vol. 8, no. 4, pp. 1252–1255, Aug. 2019, doi: 10.1109/LWC.2019.2912883.

[65] J. Chen, J. Jia, Y. Deng, X. Wang, and A. H. Aghvami, "Adaptive compressive sensing and data recovery for periodical monitoring wireless sensor networks," *Sens. Switz.*, vol. 18, no. 10, pp. 1–17, 2018, doi: 10.3390/s18103369.

[66] M. Nguyen, K. Teague, and N. Rahnavard, "CCS: Energy-Efficient Data Collection in Clustered Wireless Sensor Networks Utilizing Block-wise Compressive Sensing," *Comput. Netw.*, vol. 106, Jun. 2016, doi: 10.1016/j.comnet.2016.06.029.

[67] G. Mohandas, S. Silas, and S. Sam, "Survey on routing protocols on mobile adhoc networks," in *2013 International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s)*, Mar. 2013, pp. 514–517. doi: 10.1109/iMac4s.2013.6526467.

[68] A. B. Chandni and K. SHARMA, "Qualitative evaluation of routing protocols of MANET in wireless sensor network," *Int. J. Comput. Commun. Technol.*, 2013.

[69] A. Nagaraja, N. Mangathayaru, N. Rajashekar, and T. S. Kumar, "A survey on routing techniques for transmission of packets in networks," in *2016 International Conference on Engineering & MIS (ICEMIS)*, Agadir, Morocco, Sep. 2016, pp. 1–6. doi: 10.1109/ICEMIS.2016.7745349.

[70] K. Pang and Y. Qin, "The Comparison Study of Flat Routing and Hierarchical Routing in Ad Hoc Wireless Networks," in *2006 14th IEEE International Conference on Networks*, Singapore, 2006, pp. 1–6. doi: 10.1109/ICON.2006.302584.

[71] H. Echoukairi, K. Bourgba, and M. Ouzzif, "A Survey on Flat Routing Protocols in Wireless Sensor Networks," in *Advances in Ubiquitous Networking*, vol. 366, E. Sabir, H. Medromi, and M. Sadik, Eds. Singapore: Springer Singapore, 2016, pp. 311–324. doi: 10.1007/978-981-287-990-5_25.

[72] Z. Manap, B. M. Ali, C. K. Ng, N. K. Noordin, and A. Sali, "A Review on Hierarchical Routing Protocols for Wireless Sensor Networks," *Wirel. Pers. Commun.*, vol. 72, no. 2, pp. 1077–1104, Sep. 2013, doi: 10.1007/s11277-013-1056-5.

[73] T. Hamma, T. Katoh, B. B. Bista, and T. Takata, "An Efficient ZHLS Routing Protocol for Mobile Ad Hoc Networks," in *17th International Conference on Database and Expert Systems Applications (DEXA'06)*, Krakow, Poland, 2006, pp. 66–70. doi: 10.1109/DEXA.2006.24.

[74] P. V. Patel and B. Kadhiwala, “Broadcasting techniques for route discovery in mobile Adhoc network — A survey,” in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, Mar. 2016, pp. 671–674.

[75] J. (Terence) Chen, R. Boreli, and V. Sivaraman, “Improving the efficiency of anonymous routing for MANETs,” *Comput. Commun.*, vol. 35, no. 5, pp. 619–627, Mar. 2012, doi: 10.1016/j.comcom.2011.07.007.

[76] B. Dorronsoro, *Evolutionary algorithms for mobile ad hoc networks*. Hoboken, New Jersey: Computer society, IEEE, Wiley, 2014.

[77] K. Bayad, E. H. Bourhim, M. Rziza, and M. Oumsis, “Comparative study of topology-based routing protocols in vehicular ad hoc network using IEEE802.11p,” in *2016 International Conference on Electrical and Information Technologies (ICEIT)*, Tangiers, Morocco, May 2016, pp. 526–530. doi: 10.1109/EITech.2016.7519656.

[78] A. S. Ashoor, “Performance analysis between distance vector algorithm (DVA) & link state algorithm (LSA) for routing network,” *IJSTER*, vol. 4, no. 02, pp. 101–105, 2015.

[79] S. Iranmanesh, R. Raad, and Kwan-Wu Chin, “A novel destination-based routing protocol (DBRP) in DTNs,” in *2012 International Symposium on Communications and Information Technologies (ISCIT)*, Gold Coast, Australia, Oct. 2012, pp. 325–330. doi: 10.1109/ISCIT.2012.6380915.

[80] S. Basagni, M. Conti, S. Giordano, and I. Stojmenovic, *Mobile Ad Hoc Networking: The Cutting Edge Directions*. John Wiley & Sons, 2013.

[81] P. Singh, “Comparative study between unicast and Multicast Routing Protocols in different data rates using vanet,” Feb. 2014, pp. 278–284. doi: 10.1109/ICICICT.2014.6781293.

[82] C. Maihofer, “A survey of geocast routing protocols,” *IEEE Commun. Surv. Tutor.*, vol. 6, no. 2, pp. 32–42, 2004, doi: 10.1109/COMST.2004.5342238.

[83] P. K. Manohari and N. Ray, “Multipath routing protocols in MANETs: a study,” in *Innovation and Challenges in Cyber Security (ICICCS-INBUSH), 2016 International Conference on*, 2016, pp. 91–96.

[84] R. Kumar, S. Tripathi, and R. Agrawal, “An Analysis and Comparison of Security Protocols on Wireless Sensor Networks (WSN),” in *Design Frameworks for Wireless Networks*, vol. 82, S. K. Das, S. Samanta, N. Dey, and R. Kumar, Eds. Singapore: Springer Singapore, 2020, pp. 3–21. doi: 10.1007/978-981-13-9574-1_1.

[85] S. Abidin, V. R. Vadi, and A. Rana, “On Confidentiality, Integrity, Authenticity, and Freshness (CIAF) in WSN,” in *Advances in Computer, Communication and Computational Sciences*, vol. 1158, S. K. Bhatia, S. Tiwari, S. Ruidan, M. C. Trivedi, and K. K. Mishra, Eds. Singapore: Springer Singapore, 2021, pp. 87–97. doi: 10.1007/978-981-15-4409-5_8.

[86] H. M. A. Fahmy, “Protocol stack of WSNs,” in *Concepts, Applications, Experimentation and Analysis of Wireless Sensor Networks,* Springer, 2021, pp. 53–66.

[87] M. M. Arifeen, D. Bhakta, S. R. H. Remu, M. M. Islam, M. Mahmud, and M. S. Kaiser, “Hidden Markov model based trust management model for underwater wireless sensor networks,” in *Proceedings Of The International Conference On Computing Advancements,* 2020, pp. 1–5.

[88] P. P. Devi and B. Jaison, “Protection on wireless sensor network from clone attack using the SDN-enabled hybrid clone node detection mechanisms,” *Comput. Commun.,* vol. 152, pp. 316–322, 2020.

[89] L. Sujihelen and C. Senthilsingh, “Detecting Replica Node in Distributed Mobile Wireless Sensor Networks,” 2021.

[90] P. M. Bala, S. Usharani, and V. Abarna, “Detect the Replication Attack on Wireless sensor network by Using Intrusion Detection System,” in *Journal of Physics: Conference Series,* 2021, vol. 1717, no. 1, p. 012023.

[91] D. Hossain, Q. Mao, I. Manohar, and F. Hu, “Jamming Attacks and Countermeasures in UAV Networks,” *UAV Swarm Netw.,* pp. 207–222, 2020.

[92] M. N. U. Islam, A. Fahmin, M. S. Hossain, and M. Atiquzzaman, “Denial-of-Service Attacks on Wireless Sensor Network and Defense Techniques,” *Wirel. Pers. Commun.,* vol. 116, no. 3, pp. 1993–2021, 2021.

[93] A. Karakaya and S. Akleylek, “A survey on security threats and authentication approaches in wireless sensor networks,” in *2018 6th international symposium on digital forensic and security (ISDFS),* 2018, pp. 1–4.

[94] S. Banga, H. Arora, S. Sankhla, G. Sharma, and B. Jain, “Performance Analysis of Hello Flood Attack in WSN,” in *Proceedings of International Conference on Communication and Computational Technologies,* 2021, pp. 335–342.

[95] A. Singh, A. K. Sah, A. Singh, B. Jaint, and S. Indu, “Wormhole Attack Detection in Wireless Sensor Network Using SVM and Delay Per-hop Indication,” in *Data Engineering and Communication Technology,* Springer, 2021, pp. 39–47.

[96] N. Tamilarasi and S. G. Santhi, “Detection of Wormhole Attack and Secure Path Selection in Wireless Sensor Network,” *Wirel. Pers. Commun.,* vol. 114, no. 1, pp. 329–345, 2020.

[97] R. K. Dhanaraj, L. Krishnasamy, O. Geman, and D. R. Izdrui, “Black Hole and Sink Hole Attack Detection in Wireless Body Area Networks,” *CMC-Comput. Mater. Contin.,* vol. 68, no. 2, pp. 1949–1965, 2021.

[98] I. Kaushik and N. Sharma, “Black hole attack and its security measure in wireless sensors networks,” in *Handbook of Wireless Sensor Networks: Issues and Challenges in Current Scenario’s,* Springer, 2020, pp. 401–416.

[99] A. J. Clement Sunder and A. Shanmugam, “Black Hole Attack Detection in Healthcare Wireless Sensor Networks Using Independent Component Analysis Machine Learning Technique,” *Curr. Signal Transduct. Ther.*, vol. 15, no. 1, pp. 56–64, 2020.

[100] A. A. Abdildaeva, “Sybil Attack Detection In Wireless Sensor Networks,” in *2020 IEEE 14th International Conference on Application of Information and Communication Technologies (AICT)*, 2020, pp. 1–6.

[101] P. C. Kala, A. P. Agrawal, and R. R. Sharma, “A Novel Approach for Isolation of Sinkhole Attack in Wireless Sensor Networks,” in *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2020, pp. 163–166.

[102] A. Mousavi, M. Rezaee, and R. Ayanzadeh, “A survey on compressive sensing: Classical results and recent advancements,” *ArXiv Prepr. ArXiv190801014*, 2019.

[103] T. Wimalajeewa and P. K. Varshney, “Application of compressive sensing techniques in distributed sensor networks: A survey,” *ArXiv Prepr. ArXiv170910401*, 2017.

[104] T. Wimalajeewa and P. K. Varshney, “Compressive sensing based signal processing in wireless sensor networks: A survey,” *Online Available Httpsarxiv Orgabs170910401*, 2017.

[105] R. Middya, N. Chakravarty, and M. K. Naskar, “Compressive sensing in wireless sensor networks–a survey,” *IETE Tech. Rev.*, vol. 34, no. 6, pp. 642–654, 2017.

[106] M. Ilyas, *The Handbook of Ad Hoc Wireless Networks*. CRC Press, 2017.

[107] A. Singh, S. Sharma, and J. Singh, “Nature-inspired algorithms for Wireless Sensor Networks: A comprehensive survey,” *Comput. Sci. Rev.*, vol. 39, p. 100342, Feb. 2021, doi: 10.1016/j.cosrev.2020.100342.

[108] D. Kandris, C. Nakas, D. Vomvas, and G. Koulouras, “Applications of Wireless Sensor Networks: An Up-to-Date Survey,” *Appl. Syst. Innov.*, vol. 3, no. 1, Art. no. 1, Mar. 2020, doi: 10.3390/asi3010014.

[109] L. Gao, G. Zhang, B. Yu, Z. Qiao, and J. Wang, “Wearable human motion posture capture and medical health monitoring based on wireless sensor networks,” *Measurement*, vol. 166, p. 108252, Dec. 2020, doi: 10.1016/j.measurement.2020.108252.

[110] “MEDICAL WIRELESS SENSOR NETWORK COVERAGE AND CLINICAL APPLICATION OF MRI LIVER DISEASE DIAGNOSIS.” https://ejmcm.com/article_6867.html (accessed Jun. 26, 2021).

[111] A. Shahraki, A. Taherkordi, Ø. Haugen, and F. Eliassen, “Clustering objectives in wireless sensor networks: A survey and research direction analysis,” *Comput. Netw.*, vol. 180, p. 107376, Oct. 2020, doi: 10.1016/j.comnet.2020.107376.

[112] “A survey on compressive sensing: classical results and recent advancements.” https://jmm.guilan.ac.ir/article_4155.html (accessed Jun. 26, 2021).

[113] "Energy Efficient Data Gathering using Spatio-temporal Compressive Sensing for WSNs | Request PDF." https://www.researchgate.net/publication/348242399_Energy_Efficient_Data_Gathering_using_Spatio-temporal_Compressive_Sensing_for_WSNs (accessed Jun. 26, 2021).

[114] M. Khademi Nori and S. Sharifian, "EDMARA2: a hierarchical routing protocol for EH-WSNs," *Wirel. Netw.*, vol. 26, no. 6, pp. 4303–4317, Aug. 2020, doi: 10.1007/s11276-020-02328-w.

[115] "(PDF) Threats to Wireless Sensor Networks and Approaches to Use Homomorphic Encryption to Secure Its Data." https://www.researchgate.net/publication/346746138_Threats_to_Wireless_Sensor_Networks_and_Approaches_to_Use_Homomorphic_Encryption_to_Secure_Its_Data (accessed Jun. 26, 2021).

[116] S. Ifzarne, H. Imad, and N. Idrissi, "Secure Data Collection for Wireless Sensor Network," 2021, pp. 241–248. doi: 10.1007/978-3-030-53440-0_26.

[117] "Development of algorithms for data transmission in sensor networks based on fully homomorphic encryption using symmetric Kuznyechik algorithm - IOPscience." https://iopscience.iop.org/article/10.1088/1742-6596/1812/1/012034 (accessed Jun. 26, 2021).

[118] "Soft computing based compressive sensing techniques in signal processing: A comprehensive review." https://www.degruyter.com/document/doi/10.1515/jisys-2019-0215/html (accessed Jun. 26, 2021).

[119] "Compressive Sensing-Based Data Aggregation Approaches for Dynamic WSNs | IEEE Journals & Magazine | IEEE Xplore." https://ieeexplore.ieee.org/document/8684869 (accessed Jun. 26, 2021).

[120] P. Zhang, S. Wang, K. Guo, and J. Wang, "A secure data collection scheme based on compressive sensing in wireless sensor networks," *Ad Hoc Netw.*, vol. 70, pp. 73–84, Mar. 2018, doi: 10.1016/j.adhoc.2017.11.011.

[121] L. Kong, L. He, X.-Y. Liu, Y. Gu, M.-Y. Wu, and X. Liu, "Privacy-Preserving Compressive Sensing for Crowdsensing Based Trajectory Recovery," in *2015 IEEE 35th International Conference on Distributed Computing Systems*, Jun. 2015, pp. 31–40. doi: 10.1109/ICDCS.2015.12.

[122] "A compressive sensing–based adaptable secure data collection scheme for distributed wireless sensor networks - Zhen Liu, Yi-Liang Han, Xiao-Yuan Yang, 2019." https://journals.sagepub.com/doi/full/10.1177/1550147719856516 (accessed Jun. 26, 2021).

[123] S. Ifzarne, H. Imad, and N. Idrissi, "Homomorphic Encryption for Compressed Sensing in Wireless Sensor Networks," Oct. 2018, pp. 1–6. doi: 10.1145/3286606.3286857.

[124] P. Zhang, J. Wang, K. Guo, F. Wu, and G. Min, “Multi-functional secure data aggregation schemes for WSNs,” *Ad Hoc Netw.*, vol. 69, pp. 86–99, Feb. 2018, doi: 10.1016/j.adhoc.2017.11.004.

[125] C. Peng, M. Luo, P. Vijayakumar, D. He, O. Said, and A. Tolba, “Multi-Functional and Multi-Dimensional Secure Data Aggregation Schemes in WSNs,” *IEEE Internet Things J.*, pp. 1–1, 2021, doi: 10.1109/JIOT.2021.3077866.

[126] W. Fang, X. Wen, J. Xu, and J. Zhu, “CSDA: a novel cluster-based secure data aggregation scheme for WSNs,” *Clust. Comput.*, vol. 22, no. 3, pp. 5233–5244, May 2019, doi: 10.1007/s10586-017-1195-7.

[127] B. Alaya, L. Laouamer, and N. Msilini, “Homomorphic encryption systems statement: Trends and challenges,” *Comput. Sci. Rev.*, vol. 36, p. 100235, May 2020, doi: 10.1016/j.cosrev.2020.100235.

[128] W. Ren *et al.*, “Privacy-preserving using homomorphic encryption in Mobile IoT systems,” *Comput. Commun.*, vol. 165, pp. 105–111, Jan. 2021, doi: 10.1016/j.comcom.2020.10.022.

[129] T. Issariyakul and E. Hossain, *Introduction to Network Simulator NS2*. Springer Science & Business Media, 2011.

[130] A. S. Sethi and V. Y. Hnatyshin, *The practical OPNET user guide for computer network simulation*. Boca Raton, FL: CRC Press, 2013.

[131] L. Hogie, P. Bouvry, and F. Guinand, “An Overview of MANETs Simulation,” *Electron. Notes Theor. Comput. Sci.*, vol. 150, no. 1, pp. 81–101, Mar. 2006, doi: 10.1016/j.entcs.2005.12.025.

[132] A. Varga and R. Hornig, “An overview of the OMNeT++ simulation environment,” 2008.

[133] K. Wehrle, M. Güneş, and J. Gross, Eds., *Modeling and Tools for Network Simulation*. Heidelberg: Springer, 2010.

[134] “NS2 Book and Resources.” http://www.ece.ubc.ca/~teerawat/NS2.htm (accessed Jan. 29, 2018).

[135] Fang, W., Wen, X., Xu, J. et al. CSDA: a novel cluster-based secure data aggregation scheme for WSNs. Cluster Comput 22, 5233–5244 (2019). https://doi.org/10.1007/s10586-017-1195-7.

[136] R. Abd-Alhameed, D. Zhou, C. See, Y. Hu, and K. Horoshenkov, “Measure the range of sensor networks,” vol. 55, Jan. 2016.

[137] N. Raza, M. Umar Aftab, M. Qasim Akbar, O. Ashraf, and M. Irfan, “Mobile Ad-Hoc Networks Applications and Its Challenges,” *Commun. Netw.*, vol. 08, no. 03, pp. 131–136, 2016, doi: 10.4236/cn.2016.83013.

[138] N. Blackburn, "Rescuers turn to Israeli tech to help save trapped Thai boys," *Israel21c*. http://www.israel21c.org/rescuers-turn-to-israeli-tech-to-help-save-trapped-thai-boys/ (accessed Dec. 02, 2018).

[139] Y. Ben Chigra, A. Ghadi, and M. Bouhorma, "Taxonomy of Routing Protocols in MANETs," in *Advanced Information Technology, Services and Systems*, vol. 25, M. Ezziyyani, M. Bahaj, and F. Khoukhi, Eds. Cham: Springer International Publishing, 2018, pp. 280–288. doi: 10.1007/978-3-319-69137-4_25.

[140] G. Kaur, V. Bhatia, and D. Gupta, "Comparative Study of the performance of existing protocols of MANET with simulation and justification of an improved Routing Protocol," *They Publ. Their Res. Pap. In" Int. J. Adv. Res. Electron. Commun. Eng. IJARECE Vol.*, vol. 6, 2017.

[141] Y. B. Chigra, A. Ghadi, and M. Bouhorma, "Mobility based study of routing protocols in mobile ad hoc network," in *Proceedings of the Mediterranean Symposium on Smart City Application - SCAMS '17*, Tangier, Morocco, 2017, pp. 1–6. doi: 10.1145/3175628.3175652.

[142] C. Perkins, E. Belding-Royer, and S. Das, "Ad hoc On-Demand Distance Vector (AODV) Routing," RFC Editor, RFC3561, Jul. 2003. doi: 10.17487/rfc3561.

[143] B. Xu and Y. Li, "A novel link stability and energy aware routing with tradeoff strategy in mobile ad hoc networks," *J Commun*, vol. 9, no. 9, pp. 706–713, 2014.

[144] R. Singh and S. Gupta, *EE-AODV: Energy Efficient AODV routing protocol by Optimizing route selection process*, vol. 3. International Journal of Research in Computer and Communication Technology, 2014.

[145] S. Prasad and K. Bhatia, *RSAODV : A Route Stability Based Ad Hoc on Demand Distance Vector Routing Protocol for Mobile Ad Hoc Network*, vol. 6. 2014. doi: 10.5121/ijwmn.2014.6609.

[146] H. Dandotiya, R. Jain, and R. Bhatia, "Route Selection in MANETs by Intelligent AODV," in *2013 International Conference on Communication Systems and Network Technologies*, Gwalior, Apr. 2013, pp. 332–335. doi: 10.1109/CSNT.2013.76.

[147] P. Periyasamy and E. Karthikeyan, "Link reliable multipath routing protocol for mobile ad hoc networks," in *2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015]*, Nagercoil, India, Mar. 2015, pp. 1–7. doi: 10.1109/ICCPCT.2015.7159291.

[148] T.-C. Huang, S.-Y. Huang, and L. Tang, "AODV-Based Backup Routing Scheme in Mobile Ad Hoc Networks," in *2010 International Conference on Communications and Mobile Computing*, Shenzhen, China, Apr. 2010, pp. 254–258. doi: 10.1109/CMC.2010.313.

[149] M. Zarei, K. Faez, and J. M. Nya, "Modified Reverse AODV routing algorithm using route stability in mobile ad hoc networks," in *2008 IEEE International Multitopic Conference*, Karachi, Dec. 2008, pp. 255–259. doi: 10.1109/INMIC.2008.4777745.

[150] L. Liu, X. Li, J. Jin, Z. Huang, M. Liu, and M. Palaniswami, “Graph-Based Routing, Broadcasting and Organizing Algorithms for Ad-Hoc Networks,” in *Wireless Ad-Hoc Networks*, H. Zhou, Ed. InTech, 2012. doi: 10.5772/54146.

[151] A. Pramanik, B. Choudhury, T. S. Choudhury, W. Arif, and J. Mehedi, “Simulative study of random waypoint mobility model for mobile ad hoc networks,” in *2015 Global Conference on Communication Technologies (GCCT)*, Thuckalay, Kanya kumari district, India, Apr. 2015, pp. 112–116. doi: 10.1109/GCCT.2015.7342634.

Appendix I

CSHEAD & PC2SR users guide

The aims of this appendix is to present the environment we used to implement the CSHEAD and PC2SR schemes, and give coding details of its main modules. On the other hand, once CSHEAD and PC2SR code is finished it is fundamental to know how to integrate the new protocol into NS2 to be able to run simulations. This phase is time consuming that why we reserved section four to specify step by step the process of patching CSHEAD and PC2SR in NS2.

1. Software requirement

The development and implementation of CSHEAD and PC2SR schemes has been carried out using Ubuntu 12.04 LTS 64bit as operating system and network simulator NS 2.35. Moreover, C++ Language was used to develop the CSHEAD and PC2SR modules; TCL tool has served for establishing simulation scenarios and AWK script for post-processing the trace files generated by NS2.

2. Description of NS2's internal architecture[34]

The architecture of NS2 is based on the OSI model. The network model represents the interconnection of network elements. It consists of nodes and links. Single or multiple traffic generators, such as CBR, FTP and Telnet, can be attached to any node. In addition, the behavior

[34] Source: Technical Report No. IDSIA-24-03 « Analysis of simulation environments for mobile ad hoc networks »

of network and transport protocol is simulated by attaching the appropriate agents to the interested nodes.

The source code of NS2 is split between C++ for its core engine and OTcl for configuration and simulation scripts. The package provides a compiled class hierarchy of objects written in C++ and an interpreted class hierarchy of objects written in OTcl related to the compiled ones. The user creates new objects through the OTcl interpreter. The OTcl interpreter provides commands to create the networks topology of links and nodes and the agents associated with nodes. Implementation and simulation under NS-2 consists of four steps:

1. Implementing the protocol by adding a combination of C++ and OTcl code to NS-2's source base;
2. Describing the simulation in an OTcl script;
3. Running the simulation
4. Analyzing the generated trace files.

The implementation of a new protocol requires building the protocol procedures in C++ code and updating the OTcl configuration files to get NS2 recognize the new protocol and its default parameters.

The simulation is configured, controlled and operated through the use of interfaces provided by the OTcl class Simulator. The class provides procedures to create and manage the topology, to initialize the packet format and to choose the scheduler. The user creates the topology using OTcl through the use of the standalone classes' node and link that provide a few simple primitives.

The function of a node is to receive a packet, to examine it and map it to the relevant outgoing interfaces. A node is composed of simpler classifier objects. Each classifier in a node performs a particular function, looking at a specific portion of the packet and forwarding it to the next classifier.

Agents are another important type of components of a node: they model endpoints of the network where packets are constructed, processed or consumed. Users create new sources or sinks from the class Agent. NS currently supports various TCP agents, UDP, and other general protocols, including RTP, RTCP, and SRM.

Links are modeled either as simplex or duplex-links with a predefined capacity, delay, and queuing discipline. In addition, Links are built from a sequence of connectors' objects. The data structure representing a link is composed by a queue of connector objects, its head, the type of link, the TTL (time to live), and an object that processes link drops see figure A.1.

Various types of links are supported such as point-to-point, broadcast, and wireless. The queues are considered as part of a link. NS-2 allows the simulation of various queuing and packet scheduling such as drop-tail (FIFO) queuing, random early detection (RED) buffer management, CBQ (priority and round-robin), weighted fair queuing (WFQ), stochastic fair queuing (SFQ) and deficit round-robin (DRR).

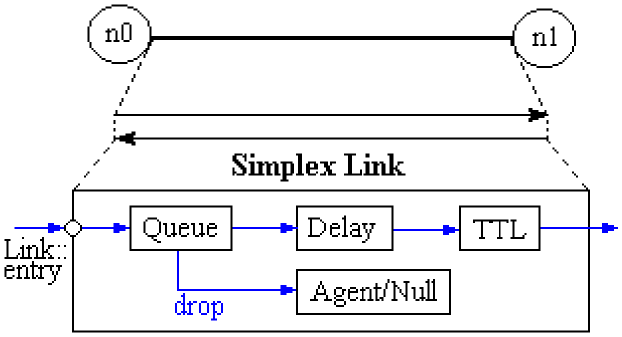


Figure A.1 Simplex link structure

The user has to specify the routing strategy (static or dynamic) and protocol to be used. Supported routing features include asymmetric routing, multi-path routing, link-state and distance vector algorithms, multicast routing, and several ad hoc algorithms.

Various types of applications can be simulated. Among them are FTP, Telnet, and HTTP, which use TCP as the underlying transport protocol, and applications requiring a constant bit rate (CBR) traffic pattern, which use the UDP transport protocol.

For the purpose of traffic generation NS-2 provides an exponential on/off distribution and it allows also to generate traffic according to a trace file.

For collecting output or trace data on a simulation NS-2 uses both traces, records of each individual packet as it arrives, departs, or is dropped at a link or queue, and monitors, record counts of various interesting quantities such as packet and byte arrivals, departures, etc., that can be associated to both packets and flows. The Network Animator (NAM) can be used for viewing NS-2 trace files for post-processing, analysis and replay of simulations.

NS2 capabilities have been extended by integration the MobileNode class, which add to the basic node object class extra functionalities related to wireless and mobile node like ability to move within a given topology, ability to receive and transmit signals to and from a network interface with an antenna, etc. The mobility features, including node movement, periodic position updates, maintaining topology boundary is implemented in C++, while plumbing of network components within Mobile Node itself (like classifiers, MAC, Channel, etc.) are implemented in OTcl.

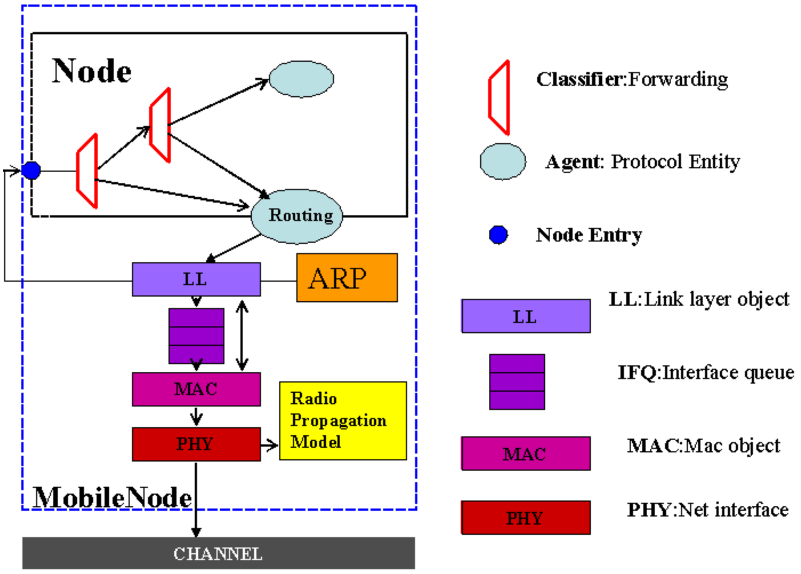


Figure A.2 Mobile node structure in NS2

The network stack for a mobile node consists of:

- Link layer,
- Address resolution protocol (ARP) module connected to the link layer,
- Interface priority queue which gives priority to routing protocol packets and supports running a filter over all packets in the queue,
- MAC layer implementing the specifications of the standard IEEE 802.11 (as well as a single-hop preamble-based TDMA MAC protocol),
- Tap Agent which receives, if allowed, all the packets from the MAC layer before address filtering is done,
- Network interface, which serves as a hardware interface used by the mobile node to access the wireless channel.

The network interface is subject to collisions and to the radio propagation model, which, in turn, receives the packets transmitted by node interfaces to their wireless channel.

The radio propagation model uses a shadowing model which, to take into account multi-path propagation effects, represents the received power in terms of a random variable. Antennas used by the mobile nodes are assumed to be omnidirectional and with unity gain.

3. How to patch CSHEAD and PC2SR in NS2

Before explaining how to patch new protocol into NS2, let's highlight the content of the main directories resulting from installing ns-allinone-2.35 package.

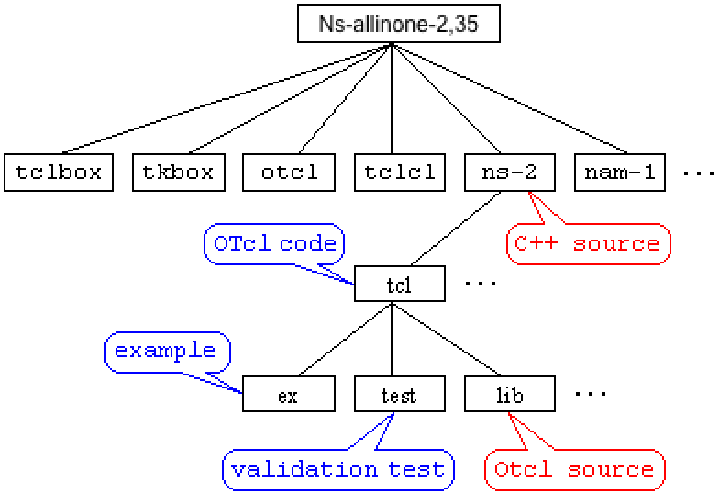


Figure A.3 NS-2 directory structure

Under ns-allinone-2.35 directory, the ns-2.35 incorporates all implementations component of simulator (either in C++ or in OTcl) such as AODV routing protocol source code directory, validation test OTcl scripts and example OTcl scripts. Within this directory, all OTcl codes are located under tcl sub-directory, and most of C++ code, which implements event scheduler and basic network component object classes.

The lib directory that belongs to tcl directory contains OTcl source codes for the most basic and essential parts of the NS implementation (agent, node, link, packet, address, routing, and etc.). The frequently used files in ns-lib.tcl, ns-packet.tcl, ns-node.tcl.

Now let's see step by step patching process of CSHEAD in NS2.35. First, prepare the cshead folder. Then execute the following steps:

- Paste ns-allinone-2.35 folder inside CSHEAD folder
- Paste mannasim folder inside CSHEAD/ns-allinone-2.35/ns-2.35 folder
- Paste packet.h,packet.cc file inside folder CSHEAD/ns-allinone-2.35/ns-2.35/common
- Paste ns-process.h file inside folder CSHEAD/ns-allinone-2.35/ns-2.35/common
- Paste channel.cc, mac-802_11.cc inside folder CSHEAD/ns-allinone-2.35/ns-2.35/mac
- Paste wireless-phy.cc,wireless-phy.h cc inside folder CSHEAD/ns-allinone-2.35/ns-2.35/mac

- Paste udp.cc file inside CSHEAD/ns-allinone-2.35/ns-2.35/apps folder
- Replace files ns-packet.tcl, ns-mobilenode.tcl, ns-lib.tcl, ns-default.tcl inside CSHEAD/ns-allinone-2.35/ns-2.35/tcl/lib folder
- In terminal, folder ns-2.35, give command ./configure
- In terminal, folder ns-2.35, give command make clean
- Paste Makefile, Makefile.in files in ns-allinone-2.35/ns-2.35 (instead of copy and paste makefile and makefile.in make some changes in existing makefile and makefile.in by copying mannasim lines after trace and paste)
- In terminal, folder ns-2.35, give command make
- In terminal, folder ns-2.35, give command make install via root user