

OUTLIER & EVENT DETECTION in WIRELESS SENSOR NETWORKS



ABDUR REHMAN

081220129

M. OMER SHABBIR

091320001

Project Supervisor : NAUMAN SHAHID

DEPARTMENT OF ELECTRICAL ENGINEERING,

SCHOOL OF SCIENCE AND TECHNOLOGY,

**UNIVERSITY OF MANAGEMENT AND TECHNOLOGY
LAHORE**

MAY, 2013

ACKNOWLEDGMENTS

In the name of Almighty Allah, the most beneficent the most merciful Who has always guided us. With His blessings this work wouldn't have been completed. We are thankful to our parents and teachers who did so much effort to see us as their successful children, today.

We would like to express our profound sense of thankfulness and pay our gratitude to our respected final year project advisor *Mr. Nauman Shahid* who was always there for us whenever we needed him. Without his guidance and support our project would have never been completed, successfully. He assisted in every possible manner and helped us to practically understand the things in comprehensive manner.

In the end we would like to express our respect to all those personals that has helped us in completing our final year project successfully.

UNDERTAKING

I certify that research work titled “**OUTLIER & EVENT DETECTION in WIRELESS SENSOR NETWORKS**” is my own work. The work has not, in whole or in part, been presented elsewhere for assessment. Where material has been used from other sources it has been properly acknowledged/ referred.

Abdur Rehman

ABDUR REHMAN (081220129)

M Omer Shabbir

M. OMER SHABBIR (091320001)

ABSTRACT

In the area of Wireless Sensor Networks (WSNs), there are measurements (readings) which deviate from the normal pattern of sensed data; caused by errors and other malicious attacks on the network. There is a requirement of such procedures which can help us identify and diagnose such data bases. Such theoretical techniques have been made. The manipulation of these theories, in Wireless Sensor Networks, is the task which is important and serious. To get an effective and efficient approach to get whole knowledge of different attributes while data sensing. Eventually, we will have the knowledge about outliers i.e. anomalies and actual events occurred. We implement few techniques and finally make a new approach to attain our goals.

TABLE OF CONTENTS

TABLE OF FIGURES	9
CHAPTER 0	11
INTRODUCTION	
<ul style="list-style-type: none">• Background• Motivation	
CHAPTER 1	
PROBLEM UNDERSTANDING and STATEMENT	12
1.1 PROJECT STATEMENT	
1.2 PROJECT DESCRIPTION	
1.3 PROBLEM UNDERSTANDING	
1.3.1 WIRELESS SENSOR NETWORK (WSN)	13
<ul style="list-style-type: none">• Wireless• Wireless Communication• Sensor• Network• Sensor Network• Wireless Network• Wireless Sensor Network<ul style="list-style-type: none">○ Architecture (Model, View, Controller, Data Base and Reading, Logic and Presentation Layers)○ Processor (Microcontroller)○ Transmission Medium○ Power Source○ Attacks	

1.3.2 OUTLIER & EVENT	24
• Outlier	
• Event	
• Outlier vs. Event	
1.3.3 MOTIVATION to OUTLIER DETECTION	27
1.3.4 CHALLENGES in DETECTION	27
1.3.5 Classification Criteria Of Detection Techniques	29
1.3.6 OUTLIER DETECTION TECHNIQUES	31

CHAPTER 2	
OUTLIER DETECTION TECHNIQUES	33
2.1 OVERVIEW	
2.2 Outlier Detection Approaches:	33
• Clustering-Based Approaches	
• Nearest Neighbor-Based Approaches	
• Statistical-Based Approaches	
• Spectral Decomposition-Based Approaches	
• Classification-Based Approaches	
2.3 COMPARISON OF Outlier Detection Techniques	35

CHAPTER 3	
THE CLUSTERING-BASED APPROACH	
3.1 INTRODUCTION	36
3.2 DIFFERENT APPROCHES	36
3.2.1 Distributed Anomaly Detection in WSN through Clustering	
3.2.2 Anomaly Detection in WSN by Clustering Ellipsoids	
3.5 GENERAL COMPARISON	39

CHAPTER 4

DISTRIBUTED ANOMALY DETECTION in WSN through CLUSTERING and MERGING

4.1	INTRODUCTION	40
4.2	METHODOLOGY	41
	4.2.1 Data Conditioning	
	4.2.2 Cluster Formation	
	4.2.3 Cluster Merging	
	4.2.3 Anomaly Detection	
4.3	Evaluating	43
	4.3.1 Detection Rate	
	4.3.2 False Positive Rate	

CHAPTER 5

ANOMALY DETECTION in WSN by CLUSTERING ELLIPSOIDS

5.1	INTRODUCTION	45
5.2	METHODOLOGY	45
	5.2.1 Measurements for Elliptical Model	
	5.2.2 Ellipsoids Clustering	
	5.2.3 Similarities	
	5.2.4 Anomaly Detection	
5.3	Evaluating	47
	5.3.1 Detection Rate	
	5.3.2 False Positive Rate	

CHAPTER 6

PRACTICAL IMPLEMENTATION: GRAPHICAL USER INTERFACE

6.1	SOFTWARE USED	49
6.2	WHAT IS MATLAB?	50
6.3	THE MATLAB SYSTEM	51

CHAPTER 7

Outlier & Event Detection in WSNs: A NEW APPROACH

7.1	INTRODUCTION	52
7.2	METHODOLOGY	52
	7.2.1 Measurements for Elliptical Model	
	7.2.2 Ellipsoids Clustering	
	7.2.3 Similarities	
	7.2.4 Anomaly Detection	
7.3	Evaluating	
	7.3.1 Detection Rate	54
	7.3.2 False Positive Rate	55
	CONCLUSION	56
	REFERENCES	57

TABLE OF FIGURES

CHAPTER 1

PROBLEM UNDERSTANDING AND STATEMENT

Figure1: Wireless Logo	13
Figure2: Wireless Communication Devices	13
Figure 3: Temperature and Humidity Sensors	14
Figure3: Sensor Node Network	15
Figure4a: MEMS technology	16
Figure5: Wireless Sensor Insight	17
Figure6: Mapping New Testament Social Networks: A Wireless Connected Network	17
Figure7: A Wireless Sensors Network	19
Figure 7.1: A General Model Representation	19
Figure7.2: A Microcontroller Chip	20
Figure7.4: Sensors with Black colored battery source in them!	21
Figure7.5: Network Attack; Eavesdropping	23
Figure1.3.2.1 (a): Outlier: between 9 and 10	24
Figure1.3.2.1 (b): Outliers can also occur when comparing relationships between two sets of data. Outliers of this type can be easily identified on a scatter diagram.	25
Figure1.3.2.3: An actual Event occurrence as an Outlier	27
Figure1.3.6: Types of Outlier Detection Techniques	32

CHAPTER 3 THE CLUSTERING-BASED APPROACH Figure2.3: Anomaly Detection Taxonomy	37
CHAPTER 3 THE CLUSTERING-BASED APPROACH Figure3.2.1: the Red points outside the circle; are the Outliers	38
CHAPTER 4 DISTRIBUTED ANOMALY DETECTION IN WSN THROUGH CLUSTERING AND MERGING Figure4.3.2: Detection Rate and False Positive Rate along with the Clusters Plot	44
CHAPTER 5 ANOMALY DETECTION IN WSN BY CLUSTERING ELLIPSOIDS Figure5.3.2: Detection Rate and False Positive Rate	48
CHAPTER 6 PRACTICAL IMPLEMENTATION: GRAPHICAL USER INTERFACE Figure6.1: MATLAB R2012a	49
CHAPTER 7 Outlier & Event Detection in WSNs: A NEW APPROCH Figure7.3.2: Detection Rate and False Positive Rate along with the event cluster plot	55

INTRODUCTION

CHAPTER

00

BACKGROUND:

For a better understanding of the background of this project, one should refer to the terms; wireless sensor networks, research surveys and implementation of different techniques for the betterment of functionality of a concept. So while talking about Wireless Sensor Networks; there was a need to have an effective and efficient way to manipulate data and find out different attributes: to get information about an occurring outlier (event, error, etc). Furthermore, to consider all possible options i.e. previous works to create a new more sound improvement and a way to handle the problem effectively.

Several real life applications of outlier detection in WSNs include; environmental and habitual monitoring, health and medical monitoring, industrial monitoring, target tracking, surveillance monitoring or any malicious attack detection. Although sensors have been used for a long time and many times these sensors have been used to create a network, new technologies developed in the last ten years have created a new vibrancy in the field of sensor networks. Interesting are the developments of technologies and protocols that have helped us to create a marriage between digital and analog sensors to work together and exchange data across packet networks. This has resulted into a multitude of applications.

MOTIVATION:

Wireless Sensor Networks recognized as WSNs, have a vital role in almost many sensitive and progressive fields today. While there are the data collecting and attributes recognition involved, there are many things to be calculated to get defined and useful information which can lead us to the improvement and also give us opportunity to have a safeguard before further damage or any abrupt hazard. For this some work has been done, which motivated us to define and implement different approaches for such cases. And it also innovate us to come up with an improvised version/technique to achieve the respective goals.

PROJECT STATEMENT AND PROBLEM UNDERSTANDING

CHAPTER

01

1.1 PROJECT STATEMENT:-

OUTLIER and EVENT DETECTION in WIRELESS SENSOR NETWORKS

1.2 PROJECT DESCRIPTION:-

Talking about the subject in WSNs; study those measurements that significantly deviate from the normal pattern of sensed data, caused by noise and errors or other malicious attacks. Traditional outlier detection techniques are not directly applicable to wireless sensor networks. This is due to the nature of sensor data and specific limitations of/in the wireless sensor networks.

Identifying misbehaviors is the most important challenge for monitoring, fault and intrusion diagnose in WSNs. A key problem can be defined as: how to minimize the communication overhead and energy consumption in the network when identifying such misbehaviors.

In the procedure to evaluate our schemes, we implemented our algorithm on the sensor data gathered from the Great Duck Island project. We present a distributed algorithm for anomaly detection in wireless sensor networks, which reduces the amount of data that needs to be communicated through the network. Then the demonstration of the approach can achieve greater accuracy in non-homogeneous sensing environments than existing methods, while achieving low communication and computational overhead in the network. A major challenge for the management of low-cost sensor networks is how to ensure the integrity of the data collected, and how to detect unusual events.

1.3 PROBLEM UNDERSTANDING

To understand the concepts of some topics, that will be leading us to an explanation of my work and methodology. So we have to learn about some of these important concepts which are described as following:-

1.3.1 WIRELESS SENSOR NETWORK (WSN)

The things which come in the way of making a Wireless Sensor Network and to understand it are as follows:

- **Wireless**

Without wire, receiving and sending as in sharing or broadcasting for example, a radio channel or sharing information without a physical medium^{[1] [2]}.

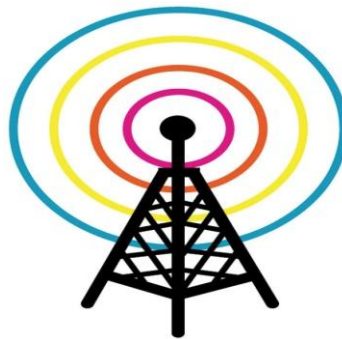


Figure1: Wireless Logo

- **Wireless Communication**

Wireless communication is the transfer of information between two or more points that are not connected by an electrical conductor. Wireless^[1] operations permit services, such as long-range communications^[3]. Information is transferred in this manner over both short and long distances.



Figure4: Wireless Communication Devices ^[7]

- **Sensor**

A sensor; also referred as detector, can be referred as a converter; a change physical quantity is measured and then the sensor converts it into a signal which can be read by an observer or by an (electronic) instrument.

Sensors have been used in both mechanical and electrical systems for a long time. It is a device that has a sensation toward a physical quantity. In short, it's a device that responds to a physical stimulus; vulnerable to faults and malicious attacks.

A good sensor obeys these rules:

- Is sensitive to the measured property only
- Is insensitive to any other property likely to be encountered in its application
- Does not influence the measured property

A sensor^{[4][5]} is a transducer which transforms a physical process into an electrical signal, which can be measured by a digital processor. Many sensors can provide information of interest for different surveillances, such as temperature, humidity, pollutant, vibration, lighting condition and magnetic sensors.



Figure5: Temperature and Humidity Sensors

- **Sensor Nodes**

The sensor nodes^[9] are equipped with sensing, processing and wireless communication capabilities. Each node is usually have a wireless radio transceiver, a small microcontroller, a power source and multi-type sensors such as temperature, humidity, light, heat, vibration, etc. The use of low cost sensor hardware can lead to a large number of anomalous measurements.

Sensor nodes communicate^[6] among themselves using wireless communication, to sense the physical world. It combines distributed sensing, computation and wireless communication technologies ^[10].

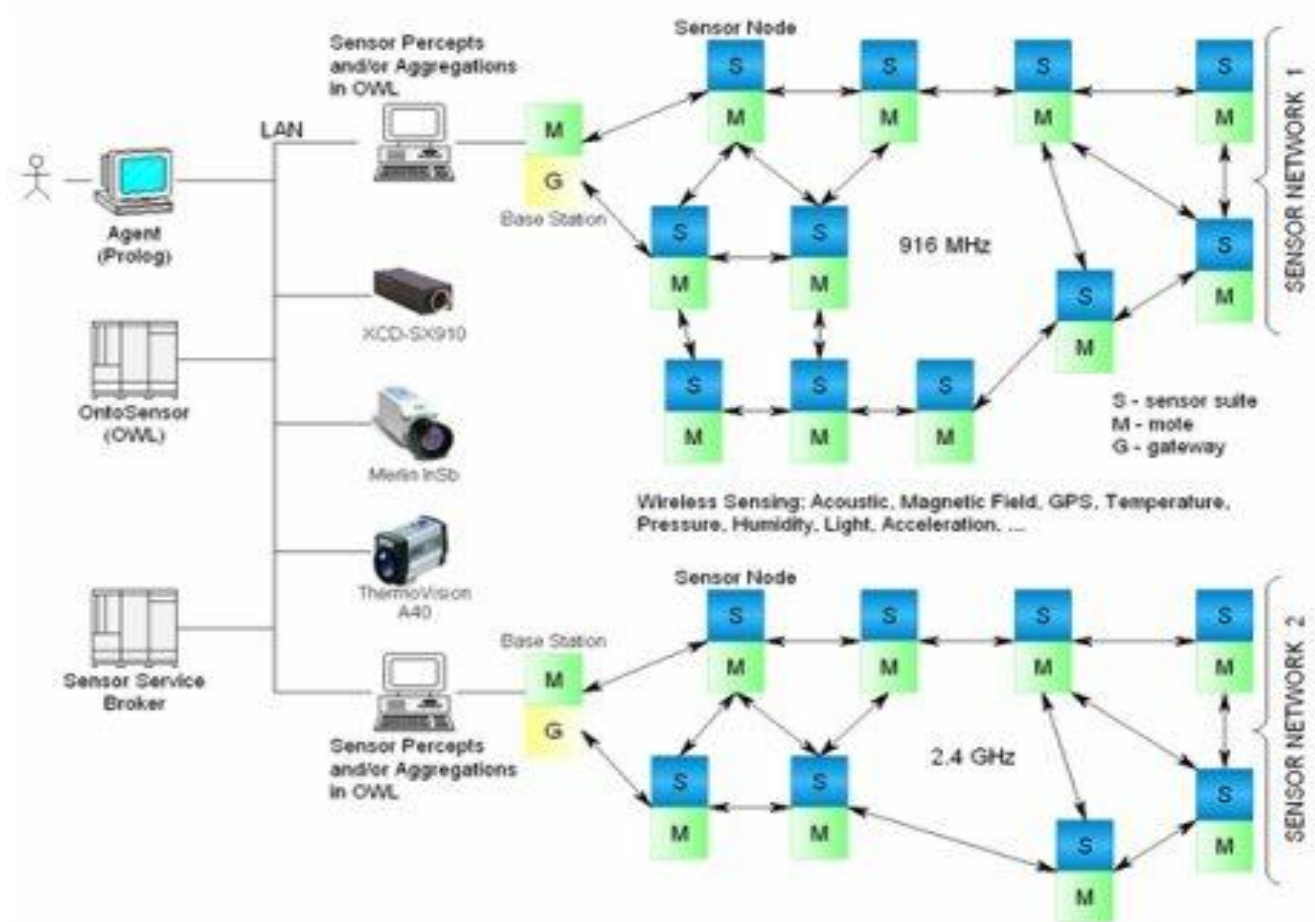


Figure6: Sensor Node Network ^[12]

Thanks to advances in MEMS (Micro-Electro-Mechanical System) technology^[8], it is now possible to integrate many sensors into a single, small integrated circuit board (the size of a quarter).

It is with very low power consumption and at a relative low cost.

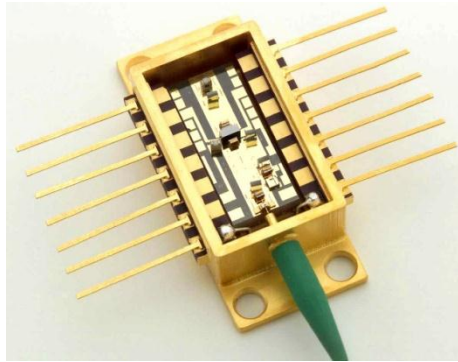


Figure4a: MEMS technology

In Figure1.3.1.4a: Alignment structures, less than a 10th of an inch in size in this optical module, make use of Sandia National Laboratories' expertise in the LIGA micro fabrication process. This photonic subsystem, by Sandia's licensing partner Axsun Technologies of Billerica, Mass., sits on an electronic assembly that occupies about 60 percent of the area of a business card (2.5 inches long x 1.7 inches wide)^[8]

- **Wireless Sensor**

Wireless sensors^{[10][11]} are the standard measurement tools equipped with transmitters to convert signals from “process control instruments” into a radio transmission. Sense Node ^[14] - Wireless Sensor Node is a low cost, rapidly deployable node used for intrusion detection purposes specifically in border and facility surveillance systems and many other important scenarios. On next page a picture of a wireless sensor with open circuit show's its insight in Figure5.



Figure5: Wireless Sensor Insight [15]

- **Network**

A system of elements; related or connected. Network can be a complex closed linking group of broadcasting systems or data bases that share information within. As an example below in Figure 8, how many places are interconnected and are sharing information. [13]

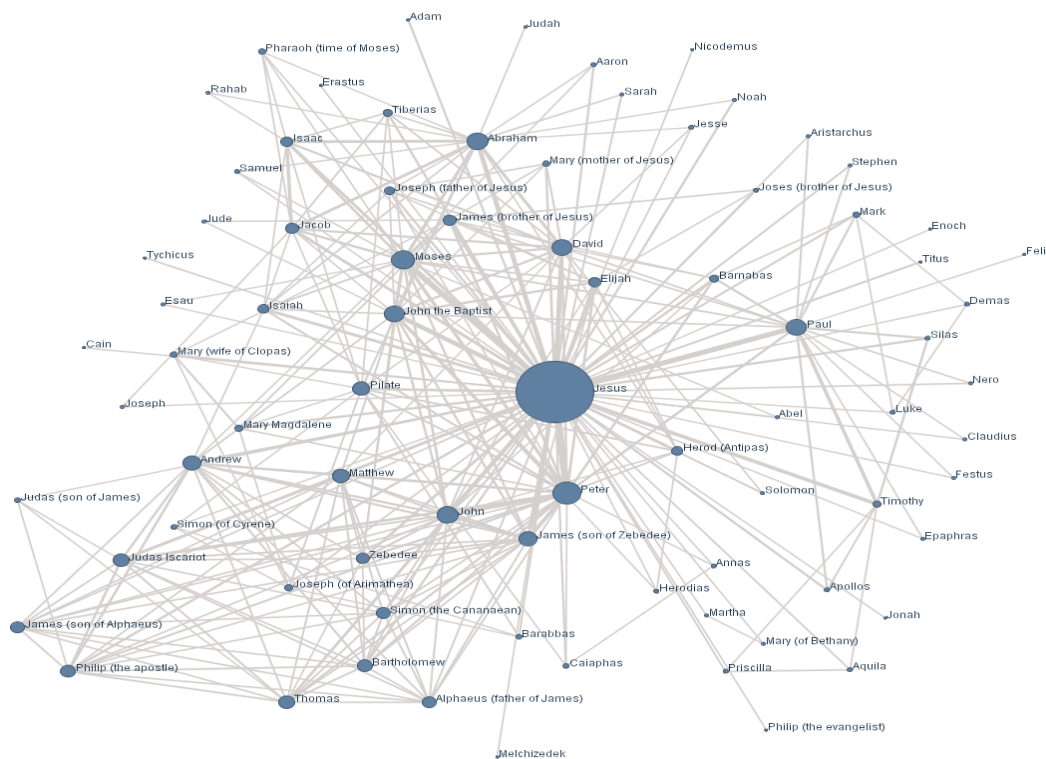


Figure6: Mapping New Testament Social Networks: A Wireless Connected Network

- **Wireless sensor network (WSN)**

Wireless sensor network (WSN); typically comprises of many tiny, low-cost sensor nodes with limited bandwidth, memory, power and capabilities to compute data. A WSN [17] is distributed over a large area, with some powerful sink nodes which gather readings of sensor nodes. The sensor nodes are equipped with capabilities like; sensing, processing and wireless communication. Each node is equipped with a wireless radio transceiver, a small microcontroller, a power source and many types of sensors such as temperature, humidity, light, heat, pressure, sound, vibration, etc. In the recent past, Wireless Sensor Technology has come up as an advanced autonomous monitoring technology and promises to be the next research frontier for many enthusiasts.

Over time, WSNs have been used for a multitude of applications. Extensive work has been dedicated for various applications of WSNs. WSNs provide a cost-effective platform for monitoring environments where the deployment of wired sensing infrastructure is too impractical or expensive. The WSN is not only used to provide fine-grained real-time data about the physical world but also to detect time-critical events.

A wide variety of applications of WSNs includes: personal, industrial, business, and military domains, such as environmental and habitat monitoring, object and inventory tracking, medical monitoring, battlefield observation, industrial safety and control, to name but a few.

These tiny sensor nodes, which consist of sensing, data processing, and communicating components, leverage the idea of sensor networks based on collaborative effort of a large number of nodes. These networks can use several different wireless technologies including IEEE802.11 wireless LANs, Bluetooth and radio frequency identification (RFID).

WSNs advances in silicon radio chips, coupled with cleverly crafted routing algorithms and network software are promising to eliminate those wires and their installation and maintenance costs. Wireless sensor networks are 'future proof'. For a simple, short-range wireless networks whose; radio components could run several years on a single battery.

Eventually, the described features ensure a wide range of different features in a sensor network that includes military, health and environmental applications, etc.

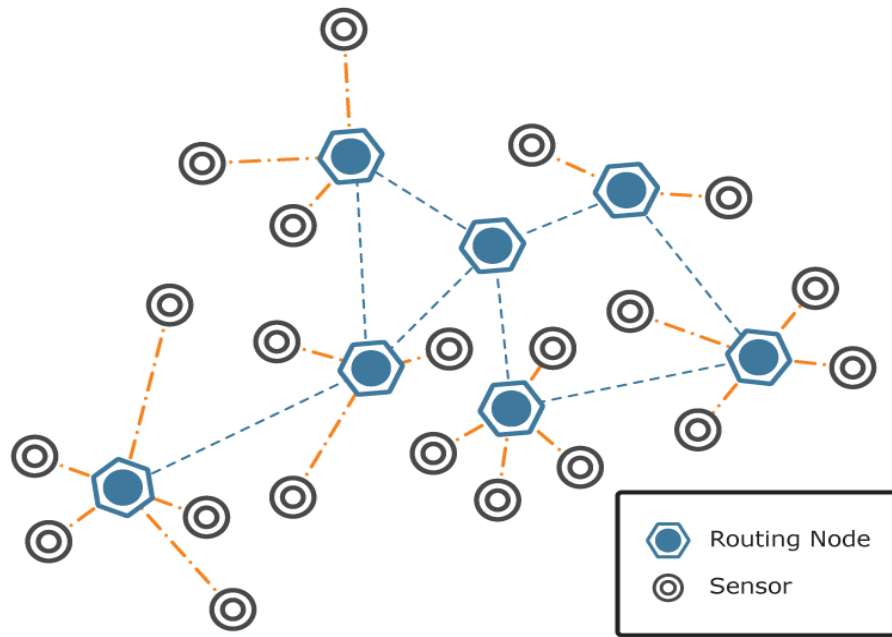


Figure7: A WIRELESS SENSORS NETWORK [18]

○ **Architecture**

Given below is a brief design and full description of the architecture of WSN:

Model: This will be more like the domain-specific representation of the information on which the applications will operate.

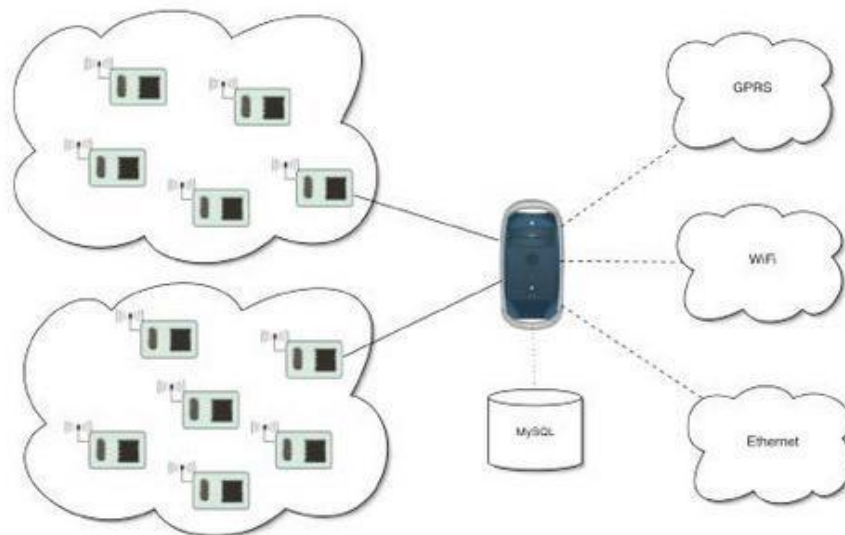


Figure 7.1: A General Model Representation

View: To make the model able for interaction; typically a user interface element.

Controller: This responds to events such as typical user actions, and invokes changes on the model.

Data Reading and Event Layer: This layer is responsible for communicating with the sensors in the field and collecting data from the base stations. This layer also validates if an event has occurred using an algorithm. If an event has occurred or not, it will be decided by admin controlled.

Database: It stores all the readings for analysis. It also stores the information about the authorized users. Storing of values allows the user to review the node behavior in the past and help in analyzing the problem in better detail.

Presentation Layer: This layer is responsible for generating the user interface. This includes reviewing the node's statistics and viewing them in graphical aspect. Details that are graphically represented are temperature, humidity.

○ **Processor (microcontroller)**

A microcontroller is a *computer-on-a-chip* used to control electronic devices. It is a type of microprocessor ^[20] comprising of self-sufficiency and cost-effectiveness, in contrast to a general-purpose microprocessor. All the processing units, memory, analog-to-digital converter, digital I/O interface and peripherals are on a single integrated circuit.



Figure7.2: A Microcontroller Chip

○ **Transmission Media**

In a wireless sensor network, the nodes are linked by a wireless medium. The medium could be by radio like RF and Bluetooth, infrared or optical waves.

Radio: The radio plays a critical role in the lifetime of a sensor node, because the overall power consumption is dominated by the energy cost of radio communication.

Typically, more than 90% of the energy consumed in the node is accounted for/by the radio transceiver, as these settings are the important design factor in determining how many nodes are needed to meet the application's maximizing system lifetime.

○ **Power Source**

The lifetime of a WSN directly depends on its power source. The energy constraint is a dominant factor of system design trade-offs for small embedded sensor devices. The scaling down in size and cost of microcontroller and sensor has outpaced that of a battery. Most sensor networks are entirely self-organizing and operate with extremely limited energy and computational resources, because most nodes may be either in inaccessible environments.

The life of a sensor node, therefore, always will be in question and it may not be able to transmit critical data when desired. The functionality of the network, therefore, depends on the consumption rate of energy by node units. This makes the power source account for a growing portion of the cost of a WSN. This cost is further magnified by the maintenance cost of replacing or recharging the batteries on a regular basis. This implies that the choice of a power source needs to be included in the system design.

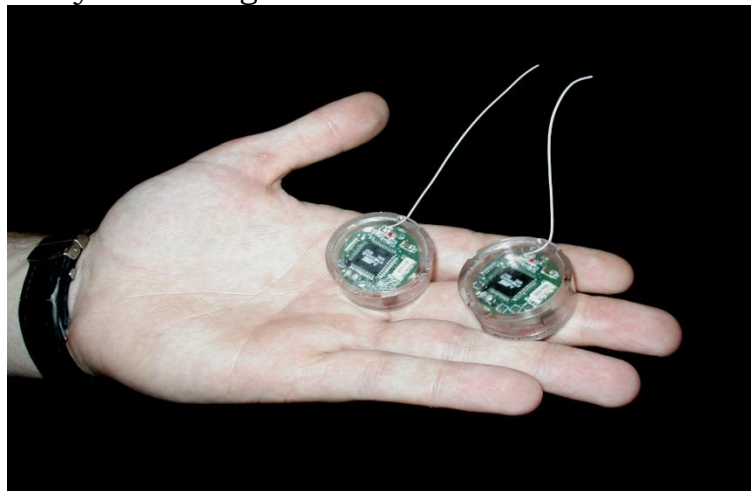


Figure7.4: Sensors with Black colored battery source in them! [21]

○ Attacks

Attacks are of several types including:

- ④ Eavesdropping
- ④ Disruption
- ④ Hijacking
- ④ Rushing

Eavesdropping: The attacker (eavesdropper) aims to determine the data that is being output by either the node or the sensor network.^[22] The attacker captures the message from the network traffic either network traffic transmitted by the nodes, or directly compromising the nodes. There are two types of eavesdropping:

- ④ Passive: Attacker's presence on the network remains unknown to the sensor nodes and uses only the broadcast medium to eavesdrop on all messages.
- ④ Active: Attacker actively attempts to discern information by sending queries to sensors, or by attacking sensor nodes.

Disruption: The intent of the attacker here is to disrupt the sensor's working. It is usually done in two ways:

- ④ *Semantically:* where the attacker injects messages, corrupts data, or changes values in order to render the aggregated data corrupt or useless.
- ④ *Physically:* where the attacker tries to upsets sensor readings by directly manipulating the environment. For example, generating heat in the vicinity of sensors will result in erroneous values being reported.

Hijacking: In this case the attacker attempts to alter the aggregated output of an application on several network sensor nodes. Someone can totally corrupt the data and can create a serious problem at the output protocol. An attacker can get a full control on the process and the information.

Rushing attack: In an on-demand protocol, a node needing a route to a destination; floods the network with *Route Request* packets in an attempt to find a route to the destination, this may cause confusions.

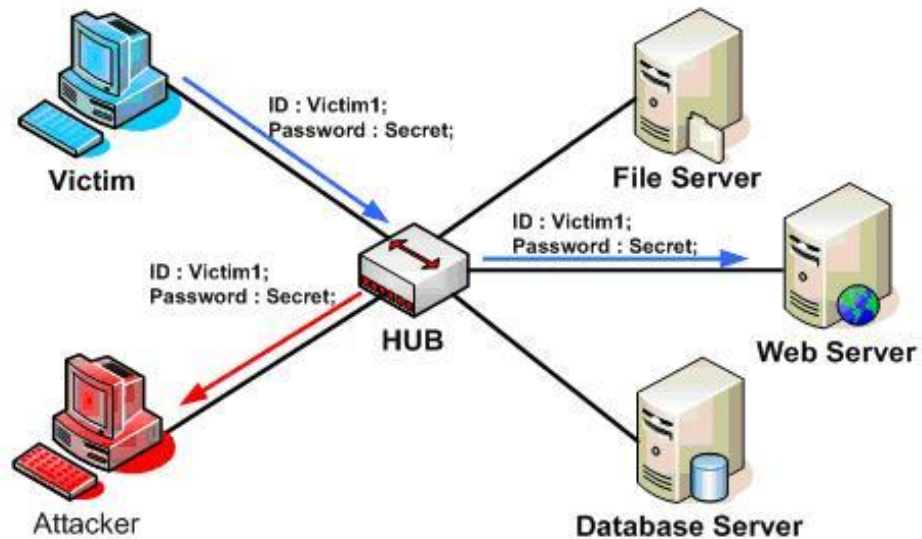


Figure7.5: Network Attack; Eavesdropping

1.3.2 OUTLIER & EVENT

The data points caused by malicious attacks are concerned with the issue of network security, but to identify them and to have a specified knowledge about them is an important goal. For resulted anomalous points from different sources, some important detection techniques are desired to specify the identity of these outliers and deal further with them. And know if an actual event is occurred or not.

Among the numerous monitoring applications of wireless sensor networks, this work will focus on outlier detection. Potential sources of outliers in data collected by WSN include noise and errors, actual events and malicious attacks. Outlier detection is also known as anomaly detection or deviation detection.

Several real life applications of outlier detection include environmental, habitual, health and medical, industrial, target tracking and surveillance monitoring. Some of the challenges of outlier detection in WSN are; high communication cost, distributed streaming data, large scale deployment, identifying outlier sources etc.

• OUTLIER

An outlier: an observation that lies outside the overall pattern of a distribution (Moore and McCabe 1999).^[24]

Usually, the presence of an outlier^[25] indicates some sort of problem. This can be a case which does not fit the model under study or an error in measurement. There are three sources of outliers in WSNs:

(1) Noise and Errors (2) Events (3) Malicious Attacks

When performing some function data, it is often best to discard outliers before computing the line of best fit. A value that "lies outside" (is much smaller or larger than) most of the other values in a set of data, as it is shown and explained in the example below:-

For example in the scores 3, 25, 27, 28, 29, 32, 33, 85, both 3 and 85 are the "outliers", which are distant from others.

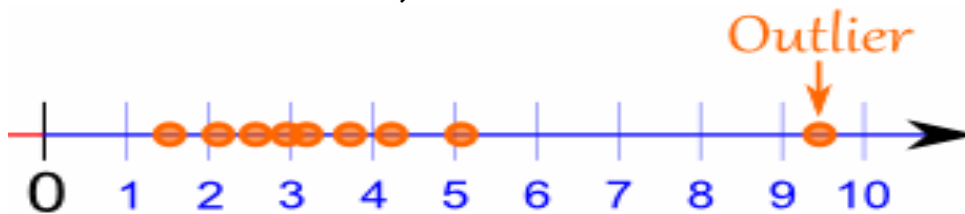


Figure1.3.2.1 (a): Outlier: between 9 and 10

As in above Fogure1.3.2.1 (a) in *Statistics* when a point in a sample is widely separated from the main cluster of points in the sample it will be declared; an **outlier**^[26], which is an observation that is numerically distant from the rest of the data.

Grubbs defined an outlier as: An outlying observation, or outlier, is one that appears to deviate markedly from other members of the sample in which it occurs.

Outliers, being the most extreme observations, may include the sample maximum or sample minimum, or both, depending on whether they are extremely high or low. It can have many anomalous causes.

A physical apparatus for taking measurements may have suffered a transient malfunction. There may have been an error in data transmission or transcription. Outliers arise due to changes in system behavior, human error, instrument error or simply through natural deviations in populations. A sample may have been contaminated with elements from outside the

population being examined. Alternatively, an outlier could be the result of a flaw in the assumed theory, calling for further investigation by the researcher.

An outlier is an observation, which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism

Barnett and Lewis: *“an outlier is an observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data.*

Potential sources of outliers in data collected by WSNs include *noise* and *errors*, *actual events*, and *malicious attacks*. Noisy data as well as erroneous data should be eliminated or corrected if possible as noise is a random error without any real significance that dramatically affects the data analysis. Outliers caused by other sources need to be identified as they may contain important information about events that are of great interest to the researchers.

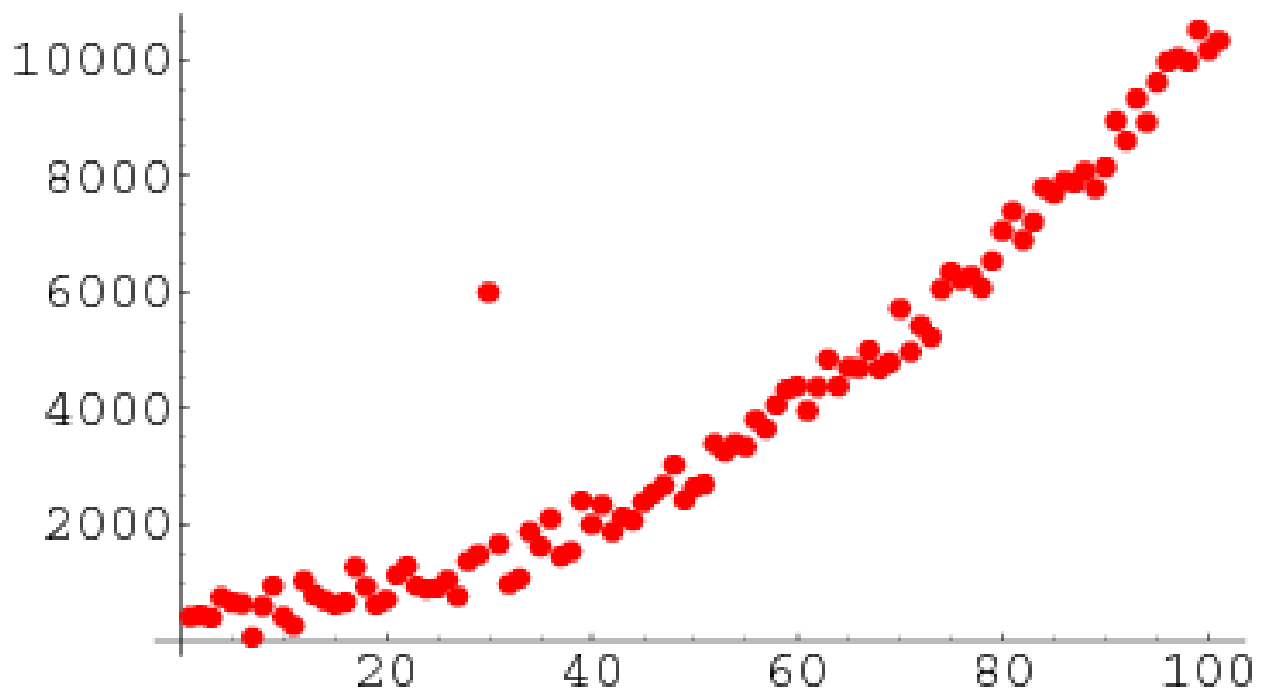


Figure 1.3.2.1 (b): Outliers can also occur when comparing relationships between two sets of data. Outliers of this type can be easily identified on a scatter diagram.

- **EVENT**

Event^[29] can be referred as something that takes place; an occurrence. A phenomenon or occurrence located at a single point in time-space, regarded as the fundamental observational entity in relativity theory.

In probability theory, an event^[30] is a set of outcomes (a subset of the sample space) to which a probability is assigned. Typically, when the sample space is finite, any subset of the sample space is an event (i.e. all elements of the power set of the sample space are defined as events). So event is just anything that takes place or happens.

In computing, an event is an action that is usually initiated outside the scope of a program and that is handled by a piece of code inside the program. An event is defined as a particular phenomenon that changes the real-world state, e.g., forest fire, chemical spill, air pollution, etc.

This ***sort of outlier normally lasts for relatively long period*** of time and changes historical pattern of sensor data. However, faulty sensors may also generate similar long segmental outliers as events and therefore it is hard to distinguish the two different outlier sources only by examining one sensing series of a node itself.

- **Outlier vs. Event**

Some of the differences between outlier and event are concluded as:

- ④ Outlier have no prior knowledge of trigger condition or semantic of any event, while event hold the trigger condition or semantic of certain event issued by the sink node.
- ④ Outlier detection aims at identifying anomalous readings by comparing sensor measurements with each other, while event detection aims at specifying a certain event by comparing sensor measurements with the trigger condition or pre-defined pattern.
- ④ To prevent normal data to be classified as outlier and thus keeping the detection rate high and false alarm rate low, while in event detection we need to prevent erroneous data which confirm to the event condition or pattern to influence reliability of the detection.
- ④ On the other hand, the common characteristic of outlier and event is that among sensor data of neighboring nodes are used to distinguish between events and errors.

Thus, outlier detection techniques are needed to make use data of neighboring nodes and spatial similarity of the sensor data. This is based on the fact that noisy measurements and sensor faults are likely to be unrelated, while event measurements are likely to be spatially correlated.

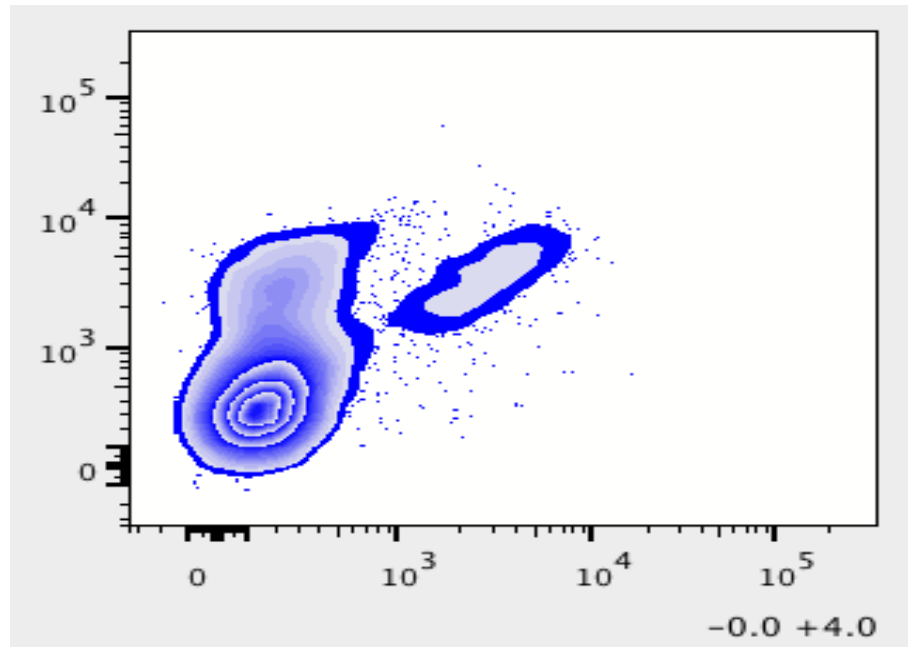


Figure1.3.2.3: An actual Event occurrence as an Outlier ^[23]

1.3.3 Motivation to Outlier Detection

As outlier detection also known as *anomaly detection* or *deviation detection*, is one of the fundamental tasks of *data mining* along with predictive modeling, cluster analysis and association analysis. Compared with these tasks, outlier detection is the closest to the initial motivation behind data mining, i.e., *mining useful and interesting information from a large amount of data*. Moreover it was a brilliant experience to develop something useful and to make whole new level for others.

1.3.4 Challenges in Outlier Detection

The context of sensor networks and the nature of sensor data make design of an appropriate outlier detection technique more challenging^[28]. According to the following reasons, outlier detection techniques might not be suitable for handing sensor data in WSNs.

- **Resource constraints**

The low cost and low quality sensor nodes have constraints in resources such as; energy, memory, computational capacity and communication bandwidth. Most of traditional outlier detection techniques have paid limited attention to reasonable availability of computational resources. They are usually computationally expensive and require much memory for data analysis and storage.

Thus, a challenge for outlier detection in WSNs is how to minimize the energy consumption while using a reasonable amount of memory for storage and computational tasks.

- **High communication cost**

In WSNs, the majority of the energy is consumed for radio communication rather than computation. For a sensor node, the communication cost is often several orders of magnitude higher than the computation cost. Thus, a challenge for outlier detection in WSNs is how to minimize the communication overhead in order to relieve the network traffic and prolong the lifetime of the network.

- **Distributed streaming data**

All of the distributed sensor data coming from many different streams may get changed. The underlying distribution of streaming data may not be known. Furthermore, direct computation of probabilities is difficult. Most of traditional outlier detection techniques that analyze data in an offline manner do not meet the requirement of handling distributed stream data. Thus, a challenge for outlier detection in WSNs is how to process distributed streaming data online.

- **Dynamic network topology**

Frequent communication failures, mobility and heterogeneity of nodes. A sensor network deployed in unattended environments over extended period of time is susceptible to dynamic network topology and frequent communication failures. Each sensor node may even be equipped with different number and types of sensors.

- **Large-scale deployment**

Deployed sensor networks can have massive size (up to hundreds or even thousands of sensor nodes). The key challenge of traditional outlier detection techniques is to maintain a high detection rate while keeping the false alarm rate low. It is a very difficult task for large-scale sensor network applications.

- **Identifying outlier sources**

The sensor network is expected to provide the raw data sensed from the physical world and also detect events occurred in the network. However, it is difficult to identify what has caused an outlier in sensor data due to the resource constraints and dynamic nature of WSNs. Traditional outlier detection technique often do not distinguish between errors and events and regard outlier as errors, which results in loss of important hidden information about events. A challenge of outlier detection in WSNs is how to identify outlier sources and make distinction between errors, events and malicious attacks.

In other words, the main question is how to process as much data as possible in a decentralized and online fashion while keeping the communication overhead, memory and computational cost low.

1.3.5 CLASSIFICATION CRITERIA OF OUTLIER DETECTION TECHNIQUES FOR WSNs

This section identifies several important aspects of outlier detection techniques specially developed for WSNs.

A. Input Sensor Data

Data which is sensed can be viewed as *data streams*, i.e., a large volume of real-valued data that is continuously collected by sensor nodes. The type of input data determines which outlier detection techniques can be used to analyze the data.

Outlier detection techniques usually consider the two following aspects of sensor data:

- 1) **Attributes:** A data measurement can be identified as outlier when its attributes have anomalous values. Most common form of data handled by anomaly detection techniques is *Record Data*:
 - a. Univariate
 - b. Multivariate

An outlier in *univariate data* with a single attribute can be easily detected if the single attribute is anomalous with respect to that attribute of other data. Thus, outlier detection techniques for WSNs should be able to analyze *multivariate data* and identify whether the attributes together display anomaly. This is simply because sometimes one of attributes individually may have an anomalous value.

2) Correlations: There are two types of dependencies at each sensor node, i.e.

- (i) Dependencies among the *attributes* of the sensor node
- (ii) Dependency of sensor node readings on history and neighboring node readings. Sensor data tends to be correlated in both time and space, especially for those data collected from environmental monitoring applications.

B. Type of Outliers

Depending on the scope of data used for outlier detection, outlier may be either *local* or *global*.

1) Local Outliers:

Due to the fact that local outliers are identified at individual sensor nodes, techniques for detecting local outliers save communication overhead and enhance the scalability. Local outlier detection can be used in many event detection applications, e.g., vehicle tracking, surveillance monitoring. Two variations for local outlier identification exist in WSNs. One is that each node identifies the anomalous values only depending on its historical values. The alternative is that in addition to its own historical readings, each sensor node collects readings of its neighboring nodes to collaboratively identify the anomalous values.

2) Global Outliers:

Global outliers are identified in a more global perspective. They are of particular interest since analysts would like to have a better understanding of overall data characteristics in WSNs. Depending on the network architecture, the identification of global outliers can be performed at different levels in the network. In a centralized architecture, all data is transmitted to the sink node for identifying outliers. This mechanism consumes much communication overhead and delays the response time. In addition, it should be mentioned

that individual nodes can identify global outliers if they have a copy of global estimator model obtained from the sink node.

C. Degree of Being an Outlier

Outlier detection techniques not only identify data that does not conform to normal pattern of sensor data, but also provide specific methods to compute the degree of which data measurements deviate from the normal pattern of sensor data.

In WSNs, outliers are measured in two scales, i.e. *scalar* and *outlier score*. Details as:

1) Scalar: The scalar scale is a zero-one classification measure, which classifies each data measurement into normal or outlier class. Thus, the output of techniques of scalar scale, which neither differentiate between different outliers nor provide a ranked list of outliers. It is a set of outliers and a set of normal measurements.

2) Outlier Score: Outlier score scale assign an outlier score to each data measurements depending on the degree of which the measurement is considered as an outlier and provide a ranked list of outliers. Such threshold is often not easy to choose and is usually user-specified and fixed. The optimal solution in WSNs is to learn the threshold and to constantly modify it with updates of arrived streaming data.

1.3.6 OUTLIER DETECTION TECHNIQUES FOR WSNs

Outlier detection in WSNs, to provide a technique-based framework to categorize current outlier detection techniques designed for WSNs. We also introduce the key characteristics and brief description of current outlier detection techniques using the proposed framework.

There are five types of outlier detection techniques^[28] for WSNs:

- Ⓢ *Statistical – Based*
- Ⓢ *Nearest Neighbor – Based*
- Ⓢ *Classification- Based*
- Ⓢ *Spectral Decomposition-based*

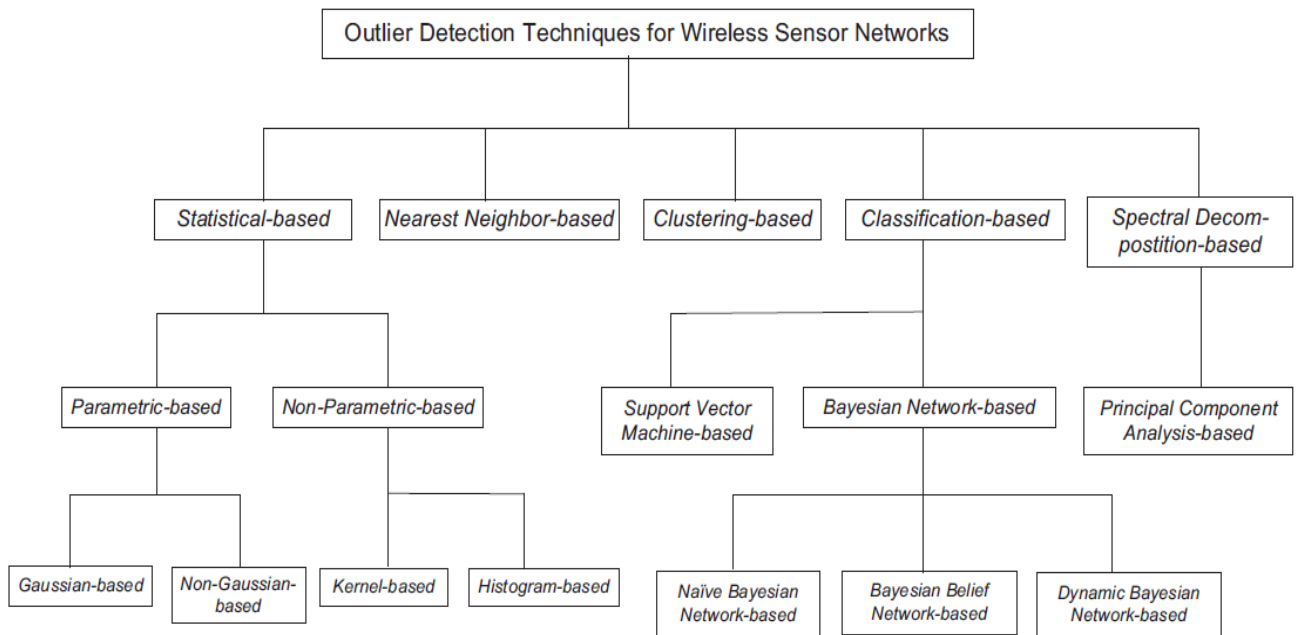


Figure 1.3.6: Types of Outlier Detection Techniques

In next chapter these approaches will be discussed to give a clear idea and brief over-views.