A HYBRID RANDOM WALK ASSISTED ZONE-BASED CLONE NODE DETECTION PROTOCOL IN STATIC WIRELESS SENSOR NETWORKS

by

MUHAMMAD NUMAN

Supervised By

Dr. FAZLI SUBHAN

Co-Supervised By

Dr. Sajjad Haider

Submitted for partial fulfillment of the requirements of the degree of MSCS to the Faculty of Engineering and Computer Science

NATIONAL UNIVERSITY OF MODERN LANGUAGES,

ISLAMABAD

June 2020

NATIONAL UNIVERSITY OF MODERN LANGUAGES

FACULTY OF ENGINEERING AND COMPUTER SCIENCE

THESIS AND DEFENSE APPROVAL FORM

The undersigned certify that they have read the following thesis, examined the defense, are satisfied with overall exam performance, and recommend the thesis to the Faculty of Engineering and Computer Sciences.

THESIS TITLE: A HYBRID RANDOM WALK ASSISTED ZONE-BASED CLONE NODE DETECTION PROTOCOL IN STATIC WIRELESS SENSOR NETWORKS

Submitted By: Muhammad Numan Registration #: 12-MS/CS/F16

Master of Science
Degree Name in Full

Computer Science
Name of Discipline

Dr. Fazli Subhan
Name of Research Supervisor Signature: _____

Dr. Sajjad Haider
Name of Co-Supervisor Signature: _____

Dr. Muhammad Akbar
Name of Dean (FE&CS) Signature: _____

Brig. Muhammad Ibrahim
Name of Director General (NUML) Signature: _____

23th June, 2020
(Date)

CANDIDATE DECLARATION

I declare that this thesis entitled "A Hybrid Random Walk assisted Zone-Based Clone Node Detection Protocol in Static Wireless Sensor Networks" is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature: _____

Name: Muhammad Numan

Date: June 23th, 2020

ABSTRACT

Wireless sensor networks (WSNs) are typically deployed in harsh and insecure environments, where the sensor nodes are generally unshielded or not tamper-resistance. As a result, WSNs suffer from such attack known as clone node or node replication attack. This attack is simply done by physically capturing the legitimate node in the network and then creating a clone with the same ID of the legitimate node. Moreover, these clones can be re-programmed for internal attacks, such as black-hole and wormhole attack, DoS attack, extract data from the network, injecting false data, disconnect the legitimate nodes through voting schemes etc. After taking control, the adversary created multiple clones on the network for various malicious activities. Therefore, clone attack is considered extremely effective because cloned devices with real information are considered real devices that can expose different protocols and sensor applications. A Systematic Literature Review (SLR) has been developed with the subject matter through which we identified the most likely solutions to encounter clone attacks are witness node based techniques. However, they have some notable weaknesses that need to be overcome and detect clones in a more effective and efficient way.

In view of the disadvantages of existing techniques based on witness nodes, this work presents a distributed technique called Hybrid Random Walk assisted Zone-Based (HRWZ) for the detection of clone nodes in static WSNs. The method is based on the Claimer-Reporter-Witness (CRW) framework. In HRWZ the network is divided into zones and a random walk approach called single stage memory random walk has been used for random selection of claimer, reporter and Zone-Leader nodes, which solved the usual problem of a simple or pure random walk. In HRWZ, Zone-Leaders are responsible for clone nodes detection locally and globally in the network. HRWZ is simulated under different settings to compare the clone detection probability, communication, memory and computation costs, with three witness node based techniques, such as RM, LSM, and RAWL. The simulation results confirm the improved performance and reliability of the proposed HRWZ technique. This scheme not only reduces the communication and storage costs, but also provides an effective method of Zone-Leader selection for high detection probability of clones.

Keywords: Wireless Sensor Network, Clone Node Detection, Systematic Literature Review, Challenges

DEDICATION

This thesis work is dedicated to my parents and my teachers throughout my education career who have not only loved me unconditionally but whose good examples have taught me to work hard for the things that I aspire to achieve.

ACKNOWLEGEMENT

First of all, I wish to express my gratitude and deep appreciation to Almighty Allah, who made this study possible and successful. This study would not be accomplished unless the honest espousal that was extended from several sources for which I would like to express my sincere thankfulness and gratitude. Yet, there were significant contributors for my attained success and I cannot forget their input, especially my research supervisors, Dr. Fazli Subhan and Dr. Sajjad Haider, who did not leave any stone unturned to guide me during my research journey.

I shall also acknowledge the extended assistance from the administrations of Department of Computer Sciences who supported me all through my research experience and simplified the challenges I faced. For all whom I did not mention but I shall not neglect their significant contribution, thanks for everything.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

WSNs	Wireless Sensor Networks
SLR	Systematic Literature Review
DoS	Denial-of-Service
HRWZ	Hybrid Random Walk assisted Zone-Based
MWSNs	Mobile Wireless Sensor Networks
BS	Base Station
RED	Randomized, Efficient and Distributed
CSI	Compressed Sensing Identification
PVM	Position Verification Method
MVP	Verification and Passing
ABCD	Area-Based Clustering Detection
LSM	Line Selected Multicast
LNCA	Local Negotiating Clustering Algorithm
SCE	Secure Cluster Election
SEC	Secure Efficient Centralized
X-RED	Extended, Randomized, Efficient Distributed
N2NB	Node to Network Broadcasting
RM	Randomized Multicast
DM	Deterministic Multicast
SDC	Single Deterministic Cell
P-MPC	Parallel Multiple Probabilistic Cells
RDE	Randomized Directed Exploration
VP	Verification Point
RAWL	Random-Walk-Based Approach

TARWL	Table-Assisted-Random-Walk Approach
NRDP	Note-Based Randomized and Distributed Protocol
DHT	Distributed Hash Table Protocol
GDL	Global Deterministic Linear Propagation
RMC	Randomized Parallel Multiple Cells Linear Propagation
ERCD	Energy Efficient Ring based Clone Detection
PAWS	Pair Access Witness Selection Technique
RE-GSASA	Residual Energy and GSA-based Simulating Annealing
CH	Cluster Head
NBDS	Neighbor-Based Clone Detection Scheme
PIV	Program Integrity Verification Protocol
TPM	Trusted Platform Module
TPIV	TPM-enabled PIV
LTBRD	Location and Trust-Based Node Replica Detection Distributed Method
Pro-Se	Proximity Service
CINORA	Cell-Based Identification of Node Replication Attacks
RQs	Research Questions
SFs	Search Filters
QA	Quality Assessment
CRW	Claimer Reporter Witness-node based technique
IoTs	Internet of Things

LIST OF SYMBOLS

N_n	Total number of nodes in the network
N_z	Number of zones in the network
Z_g	Randomly selected nodes in the zones
$N_{zl.N}$	Number of Zone_Leaders in the network
D	Number of neighbors of each node
loc_n	Node location information
F	Size of the location claim (bytes)
τ_{sr}	Transmission cost for sending and receiving one bit
τ_{sv}	Energy cost for signature check/verification
τ_{sg}	Energy cost to sign location claim
P_f	Location claim probability
P_s	Location claim storing probability
P_d	The replica/clone detection probability
λ_n	The average path length or distance between any two nodes in the network
$\lambda_{zL's}$	The average path length or distance between randomly selected Zone-Leaders
ID_n	Node Identity
K_n^{pvt}	Node n private key
K_n^{pub}	Node n public key
$Sig\ \{M\}_{K_n^{pvt}}$	The signature of node 'n' on message M
$\|$	Concatenation symbol
$ID_n, loc_n, Sig\{H(ID_n \| loc_n)\}_{K_n^{pvt}}$	Location Claim Format

Chapter 1

Introduction

Sensor nodes in Wireless Sensor Networks (WSN) gained remarkable attention for being smaller in size, cheaper and smarter. They are characterized by the fact that they are low cost, flexible, efficient, reliable and easy to be deployed. In the literature, sensor nodes in WSN are defined is a technology to sense data and perform actions with the capability of connectivity [1]. A WSN is a group of cheap/low cost sensor nodes that are lack of many resources but can cooperatively gather monitored information from the harsh and hazardous environments. These intelligent low cost sensor nodes consist of memory, a power supply, processor, as well as an actuator. Besides, they are not tamper resistant. Figure 1.1 shows the general architecture of a sensor node. According to [2] these compact, low cost, small sensor nodes really become thanks to technological advances where thousands of them could be deployed in the desired areas. These sensor nodes can collect, process and communicate data [3]. These tiny sensor nodes are mainly used to capture the physical objects, convert them to digital form and to send to a certain server (sink/base station). Not only that, but these sensor nodes can also detect things that are not physically present and possibly even invisible, such as gas and temperature. Thus, hardware manufacturing is rapidly growing and producing new sensor nodes in the market, such as Arduino, Raspberry Pi, Intel Edison, NodeMCU, Teensy, and so on.

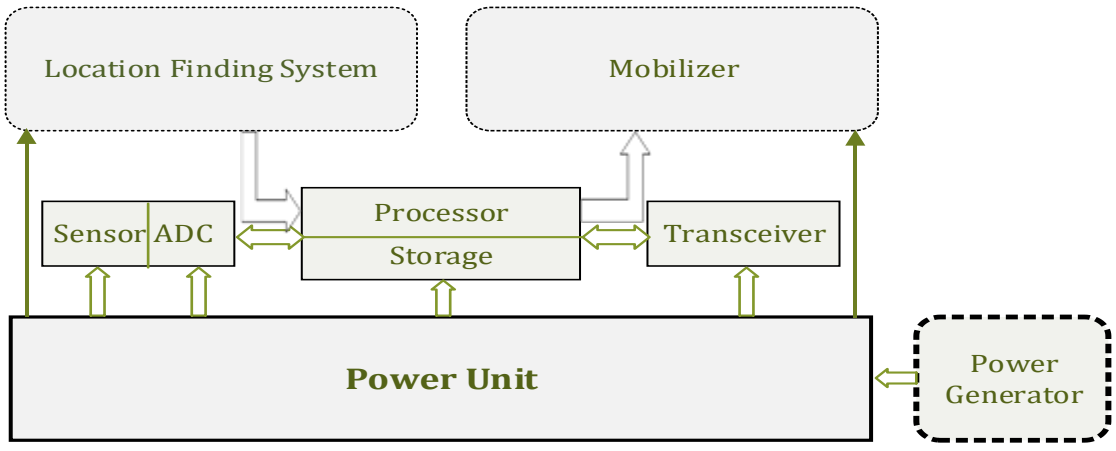


Figure 1.1: General sensor node architecture

1.1 WSNs applications

It has been found that these sensor networks could be widely used, as in military applications, environmental monitoring, commercial or any human centric applications, etc. [4], as shown in Figure 1.2. The success that these sensor nodes can achieve in military applications is based primarily on their rapid deployment, self-organizing and fault tolerance, which are very useful for military command, monitoring, control, communication, intelligence, computer and target systems. These sensor nodes could also be used in the health sectors to monitor and support disabled patients. They can also be used in other sectors and applications, such as inventory management, product quality monitoring, and disaster area monitoring [3]. According to [5] future applications would be invented using these sensor nodes to be used in more areas, such as monitoring pollution, water quality, forest fires, building security, people's heart rate, etc. The good thing about these nodes is that they can reduce the processing of raw data and turn it into useful aggregated information while being protected and completely secure.

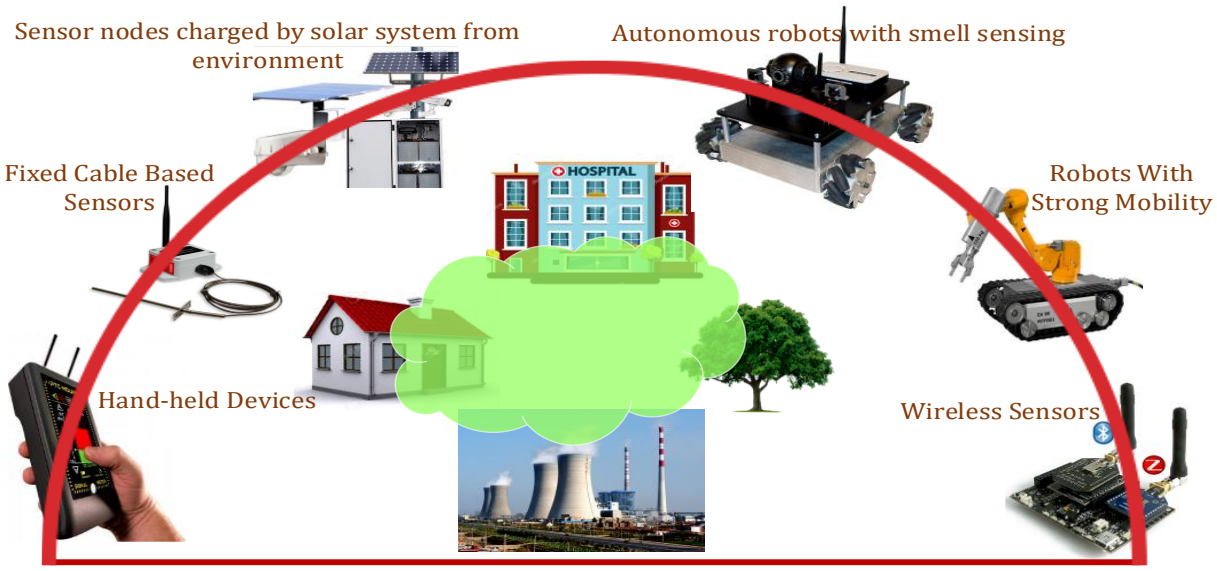


Figure 1.2: Various types of sensors monitoring environment, industries, smart homes etc.

1.2 Problems with WSNs

A very high demand and low costs do not mean that these nodes are problem-free. However, there are many problems that need to be resolved in order for these sensor nodes to be more useful, such as the node's hardware design, networking problems, and also some security

problems [1]. Besides, [5] stated that traditional security techniques that are used in the traditional devices cannot be directly used in the devices with these sensor nodes. The most posed challenges to devices using these new sensor nodes are the desire to make them cheap; thus, they're limited in energy, communication and computation capabilities. There is also the challenge of being exposed to physical attack such as node replication attack; due to being deployed in accessible areas. This is of course in addition to the security issues these sensor nodes pose for being closely interacting with physical environments and people. This is why these security problems are new, challenging and require new solutions that obviously need intense research with the matter.

1.3 WSNs Security

Sensor nodes are known to be resource constrained and lack of tamper-resistant hardware. It is therefore important to equip these sensor nodes in order to counteract the compromises [6]. Since security is considered the most important component in each and every system, it is better not to use a separate module for the provision, as explained in [5]. Where a secure system requires security to be merged into each component and never separated. This is why to reduce the risk of a physical or internal attacks, such as the privacy, secrecy, authentication, node capture attacks, signal or radio jamming, sybil attack, node outrage, sinkhole, selective forwarding attack, wormhole, or even robustness to denial-of-service attacks. Looking at to physical attacks on wireless sensor networks, it was found that the most dangerous physical attack is the node replication attack due to the poor protection of the sensor nodes. This would make it easier for an attacker to attack the nodes, replicate them in a series of clones, and finally taking them over the network. This is why, clone detection is very important in order to detect illegal copies and to protect the sensor networks where the clones have a significant impact on network routing, data collection, key distribution, etc. Therefore, networks should be secure and able to identify any threat or attack that they could be exposed to [2].

1.4 Taxonomy of attacks

As shown in Figure 1.3, sensor networks could either be threatened by layer dependent or layer independent attacks.

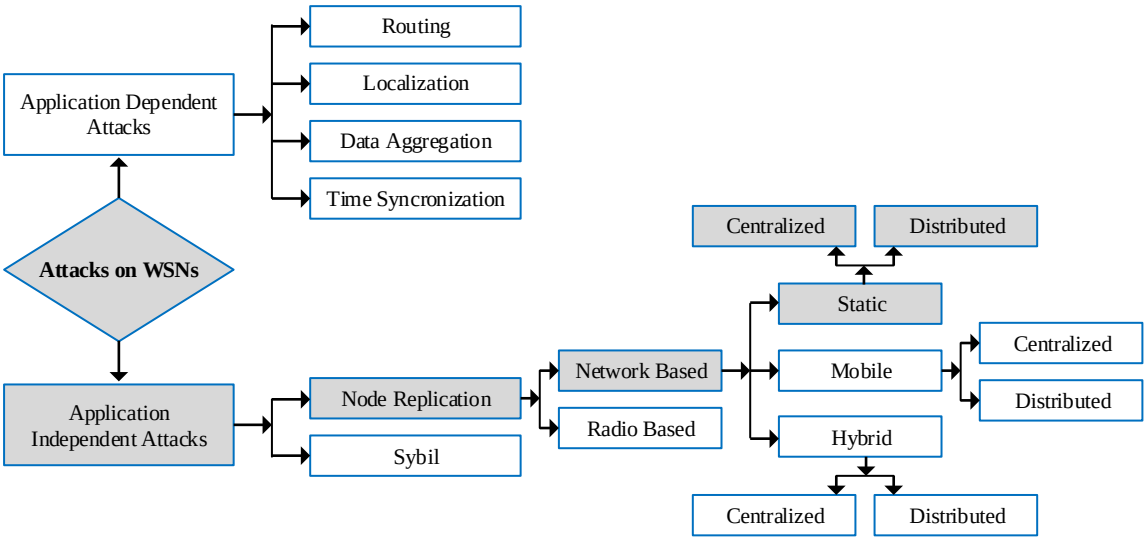


Figure 1.3: Taxonomy of attacks

1.4.1 Layer dependent attacks

Application dependent attacks are launched on different OSI layers targeting certain network functionalities, such as localization, data aggregation, routing and time synchronization. Various schemes have therefore been proposed to protect sensor networks from such attacks. For example, routing security schemes are proposed to protect sensor networks from routing interference attacks [7, 8], while authentication schemes [9-11] are proposed for mitigating false data injection attacks. Data aggregation protocols [12-15] are also proposed to eliminate the data aggregation attacks. There are also many proposals [16-22] to protect sensor networks from time synchronization and node localization attacks. The most important thing is to use attack-resistant approaches / techniques to detect and remove attacks as quickly as possible, to reduce damage costs and losses.

1.4.2 Layer independent attacks

The application-independent attacks are not launched at OSI levels. They are targeting object tracking, fire alarms, battlefield surveillance, etc. The two most dangerous attacks are: node replication attack and sybil attack [23, 24]. Node replication attacks differ from sybil attacks and can even be viewed as opposite versions of each other. A sybil attack occurs when an attacker creates multiple IDs for a single node. Sybil attack is also harmful in that it can reuse any logical ID for more than one physical sensor node to abort or damage the protocols used to combat

various sensor network attacks, i.e. data aggregation, routing, distributed storage, resource allocation, etc. Sybil attacks could be overcome by techniques and mechanisms based on RRSI [25] or by knowledge based authentication mechanisms of a fixed key set [26-30]. On the other hand, the node replication attack occurs when an attacker physically captures a node, then creates clones or replicas thereof, and ultimately deploys these clones into strategic positions on the sensor networks to form dangerous malicious activities. This type of attack could only be faced with certain global countermeasures for clone detection [9-11].

1.5 Motivation

As mentioned earlier, due to the sensitive nature, WSN can be exposed to a huge array of physical and internal attacks, as node replication attack, signal/radio jamming, denial of service (DoS) attack, node outrage, sybil attack [23, 24], sinkhole, wormhole, and selective forwarding attack [31]. This thesis focuses on node replication / clone or identity attack, which are one of the most dangerous WSN attacks. This is simply done by physically attacking, capturing a legitimate node and then creating a clone of the captured node with same ID. The attacker would then create multiple clones across the entire sensor network. In addition, some clones could even be reprogrammed to attack the sensor network internally, i.e. black hole attack, DoS attack, wormhole attack, extracting data from the network, separating the legitimate nodes through voting schemes, etc. [32]. Thus, these attacks are really harmful for WSN. This is why the need to be detected the clones as soon as possible for the adversary not to be capable of changing the protocol behavior through capturing the network security [33]. This is due to the fact that clones have logical and true identities and legitimate information that the opponents have promised. In addition, these clones cannot be distinguished from the real nodes because they also participate in network operations.

Nevertheless, this thesis aims to present a new technique apart from the traditional schemes to detect clone attack/node replication attacks in WSNs with: reduced communication, memory and storage overheads with high detection probability of clone nodes.

1.6 Problem statement

It should now be clear that WSN is resource constrained and exposed to a variety of physical and internal attacks that cause many problems. especially computing and security problems. WSN are vulnerable to a lot of security attacks, in which the most harmful attack is clone attack

or what is called the node replication attack which would later affect the network operation by internal attacks. In order to overcome the limited resources of the sensor nodes in the WSN, a lot of research work is being carried out in many areas such as low-power hardware design and routing algorithms [34], topology control and optimization issues [35, 36], etc. This is mainly to extend the network lifetime; due to the difficulty of charging or even exchanging the sensor node batteries.

Focusing on the node replication/clone attacks that are mostly due to the unshielded of sensor node hardware and deploy in hostile environment as explained earlier in the previous sections. The attacker physically attack the legitimate node first, then steals the information and the IDs and creates replicas of these legitimate nodes; which are considered for an internal attack. Then the adversary finally deploying these clone nodes in a strategic position in the network. Therefore, in order to prevent further internal attacks and changes in the protocol behavior, these clones must be detected quickly.

When searching for the main reasons why WSN sensor nodes are susceptible to all these types of attacks, it was found that these nodes turned out to be unshielded due to the desire for cost reduction and smaller in size. Therefore, there is no hardware support for manipulation security. Also the fact that the attackers are so smart that they can insert the newly created clones so easily and use information and logical identities from existing nodes to make them look legitimate and easily participate in network operations as really legitimate nodes [37]. This explains why traditional secure routing schemes [7, 8] and authentication techniques [9-11, 38-40] are not able to detect or prevent clone attack.

Many proposals have been presented for developing some techniques that might solve these issues and protecting the WSN from such attacks. However, most of them have drawbacks. The only exception is the distributed witness node based techniques that seem to have very promising results. This is why this thesis is going to use witness based techniques in which will overcome the limitations and drawbacks of these techniques as:

- The selection of the witness nodes is very sensitive and deterministic; where the adversary capture and clone the legitimate node.

- There is what is called the “uneven witness distributed problem”; where the nodes are selected and distributed unevenly over the network; causes center crowded problem. They could either be an easy target to be attacked or they could be exhausted very quickly.
- There is also a so called “tradeoff” between nodes with high detection probability and communication and memory overheads. If the latter is the preference, then the former would get weaker and the node would be vulnerable to easily attack.

1.7 Research questions

In this thesis the following questions are going to examined:

a) *How to select and secure witness-nodes (Zone-Leaders) for clone detection?*

In order to efficiently and effectively detect clone attacks in WSN, the claimer-reporter-witness based techniques should have certain characteristics. In previous study, it has been found that the witness nodes are the most important part in the claimer-reporter-witness based techniques, which help in detecting clone nodes. These are because the witness nodes are chosen randomly and not deterministically which make it difficult for the adversary to attack, compromise and replicate the nodes. However, they should be evenly distributed across the network to save their energy for long periods of time, also they will not become the focus of attention to the target of the attackers. Nonetheless, this work is concerned with the development of a technique based on distributed claimer-reporter-witness for distributed witness selection with random and non-deterministic witness node selection.

b) *How to increase clone detection probability while keeping communication and memory overheads moderate?*

In general, the number of witness nodes in a network determines the likelihood of clone detection, whether it is high or low. In the study it has been found that the problem is that as the number of witness nodes in the network increases, the likelihood of clone nodes detection increases, but this also increases the amount of communication and storage overhead. It is therefore one of the main concerns of this work to develop a new method to overcome this tradeoff between detection probability, communication and memory overheads. Not only that, but also to make it easier to detect a single clone node on the network, which is considered more difficult than detecting two or more clones in a network.

c) *What tradeoff cost a protocol would bear for applying a method of notable security and high probability for clone detection, in terms of: communication, memory and computation?*

The use of sensor node resources and the likelihood of clone detection are largely related. There is a tradeoff between to develop a witness node with higher security and clone detection ability. If the resources of sensor nodes are consumed efficiently, then the communication, computation and memory costs would remain moderate. Thus, it is another concern for this thesis.

1.8 Research aims and objectives

This research aims to develop a distributed solution for clone detection in the static WSNs with moderate overhead. This would be done by identifying the ideal requirements for designing and developing technique based on distributed witness nodes; detect clones with high probability and moderate overhead. Also the research aims to implement the proposed technique, evaluate and analyze the performance under different settings and performance metrics, then compare with existing witness node based techniques.

1.9 Scope of study

WSN could either be static or mobile in nature. The sensor nodes in the static WSN have fixed positions after being deployed. However, sensor nodes in the mobile WSN can move freely to interact with the physical environment after deployment. These mobile sensors have full autonomy and ability to compute; sense and communicate just static sensor nodes. These sensor nodes can easily communicate with each other to share or exchange information within the range. There is a difference between how data is distributed in static and mobile WSN. Not only that static and mobile WSN differ in their properties and characteristics, but also both in the clone detection techniques. With the focus on the static WSN, the goal of this MS thesis is to develop a technique where each node is deployed in a fixed position on the network to detect clone nodes when it is determined that a node is assigned to more than one location. This gives an alert that there is something wrong in the network. Furthermore, the techniques that could be used with static WSN would be inapplicable for mobile WSN.

1.10 Thesis contributions

- In this thesis work, a new distributed HRWZ solution for node replication or clone attack in the static WSN is presented to identify the ideal basis and requirements for developing the distributed claimer-reporter-witness based techniques to detect clones effectively and efficiently.
- HRWZ randomly divided the network into zones, and then select Zone-Leaders through single stage memory random walk approach to detect clones in the network. The random walk in our proposed method overcome the weaknesses of pure random walks of lower detection probability of nodes and energy consumption of sensor nodes.
- Provided more accurate and efficient results through simulation results under different settings by evaluating different metrics such as clone detection probability, communication, computation and memory overhead with the existing Claimer-Reporter-Witness based techniques RM, LSM and RAWL.

1.11 Thesis organization

This thesis is divided into six chapters, including this chapter. This chapter briefly introduces what wireless sensor networks (WSNs) really is, their components, applications, problems, different types of attacks on WSNs and security challenges. This chapter also presents the aim of the research problem, objectives, and the contribution to the problem.

In chapter two presented in detail the background of node replication or clone attack, steps of node replication attack and its effects. In this chapter we review the various mentioned clone detection mechanisms in the literature; examining their advantages and their avoidable weakness points.

Chapter three presents the Systematic Literature Review (SLR) of the study that how we investigated the clone detection schemes and challenges, also that what are the essential and ideal requirements for clone detection techniques to increase the accuracy and efficiency.

In the fourth chapter then proposed a distributed solution for node replication or clone attack in static WSNs with detailed working concepts of protocol.

Afterwards, chapter five presents the results and comparison of the proposed and selected Claimer-Report-Witness framework based techniques i.e. RM, LSM and RAWL. The results are

evaluated under different metrics such as detection probability, the communication, computation n and memory overheads.

Finally, chapter six summarized the thesis contributions, limitation and presented future direction.

1.12 Chapter summary

This chapter briefly introduces wireless sensor networks (WSNs), their components, applications, problems, different types of attacks on WSNs and security challenges. This chapter also presents the aim of the research problem, objectives, and the contribution to the problem.

Chapter 2

Literature Review

Moving from the introductory chapter to a more detailed one, this chapter aims to briefly explain clone node attach, the different steps taken for node replication attack, the consequences of the attack and how it negatively affects static WSNs security. This chapter also aims to examine the different already existing techniques for clone node detection; identified their advantages and disadvantages.

2.1 Background

2.1.1 Node replication attack or clone attack

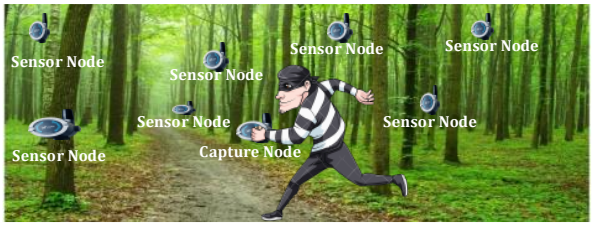
This is considered one of the most dangerous attacks on static WSNs that threaten their security, which is mainly due to the fact that the sensor nodes are unshielded. This is a physical attack; where the adversary first targets a legitimate node to capture, then he/she uses its identity to create clones to be deployed throughout the network that causes for internal attacks such as the privacy, secrecy, authentication, node capture attacks, signal or radio jamming, sybil attack, node outrage, sinkhole, selective forwarding attack, wormhole, or even robustness to denial-of-service attacks.

The adversary could also disconnect the captured legitimate node from the network and cancel the node revocation protocols. This might, then, lead to a changing in the protocol behavior, as well as the security system of the network [33]. Thus, clones should be detected as quickly as possible to prevent further damages; because they actually have legitimate IDs, information and cryptographic materials before being deployed in the network. Because, these clones are operating the same way as legitimate nodes.

2.1.2 Steps of node replication attack

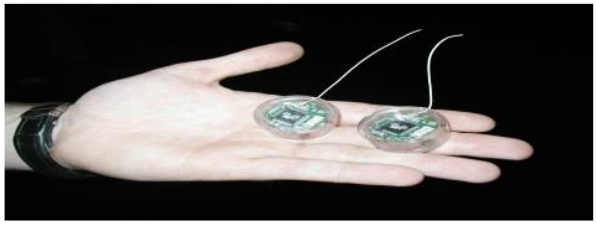
It has been agreed upon the fact that sensor nodes in static WSNs are unshielded or not tamper resistant. This is considered one of the major reasons why WSNs are unsecure and vulnerable to physical attack, such as node replication attack. The entire process of this type of attack takes place in four main steps, which are shown in Figure 2.1.

Sensor nodes deployed in the field

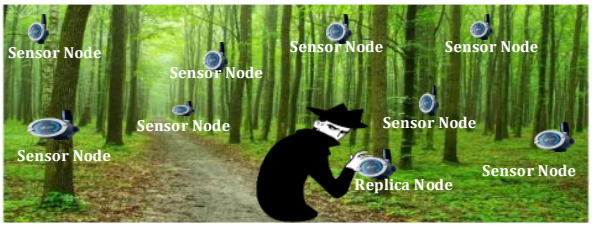


Physically capture the node

Collect all the secret informations



Replicate of the captured node



Deploy the replica node in interest area

Figure 2.1: Node capture attack

Looking at the steps in Figure 2.1. First and foremost, the adversary who physically captures the legitimate node on the network. Then he / she collects all of the secret credentials (i.e., IDs, information, data, etc.) of the legitimate nodes. He then uses all of the information gathered to create clones and ultimately deploys them in a strategic position on the network. With the help of these clones, the opponent can carry out malicious activities or carry out more internal attacks on the network or even stop the operation of the legitimate nodes and monitor and control all communication within the network.

The entire procedure for starting and detecting clone attacks is shown in the flowchart shown in Figure 2.2. The flowchart shows everything from the first stage of physical capturing of legitimate node. If the absence or tamper resistance of the node is detected, then prevent the node from operation in the network. If not detected then, stole / extract the secret credentials of the legitimate node captured in the second phase. To reprogram the legitimate node and new hardware to be used is presented in the third stage. In the fourth stage the clone has been created with deploying fake ID and information for a purpose. In the fifth stage the adversary then deploying these clones into the strategic position on the network to monitor or perform insider

attacks. The flow chart presents how these clones are being detected and how they could be prevented from causing further damages to the network using different techniques in the sixth stage.

If the attack is detected from the first stage, there would be no clone attack. However, if the clone attack took place, it can still be detected using many techniques and schemes, but of course the network would do great damage.

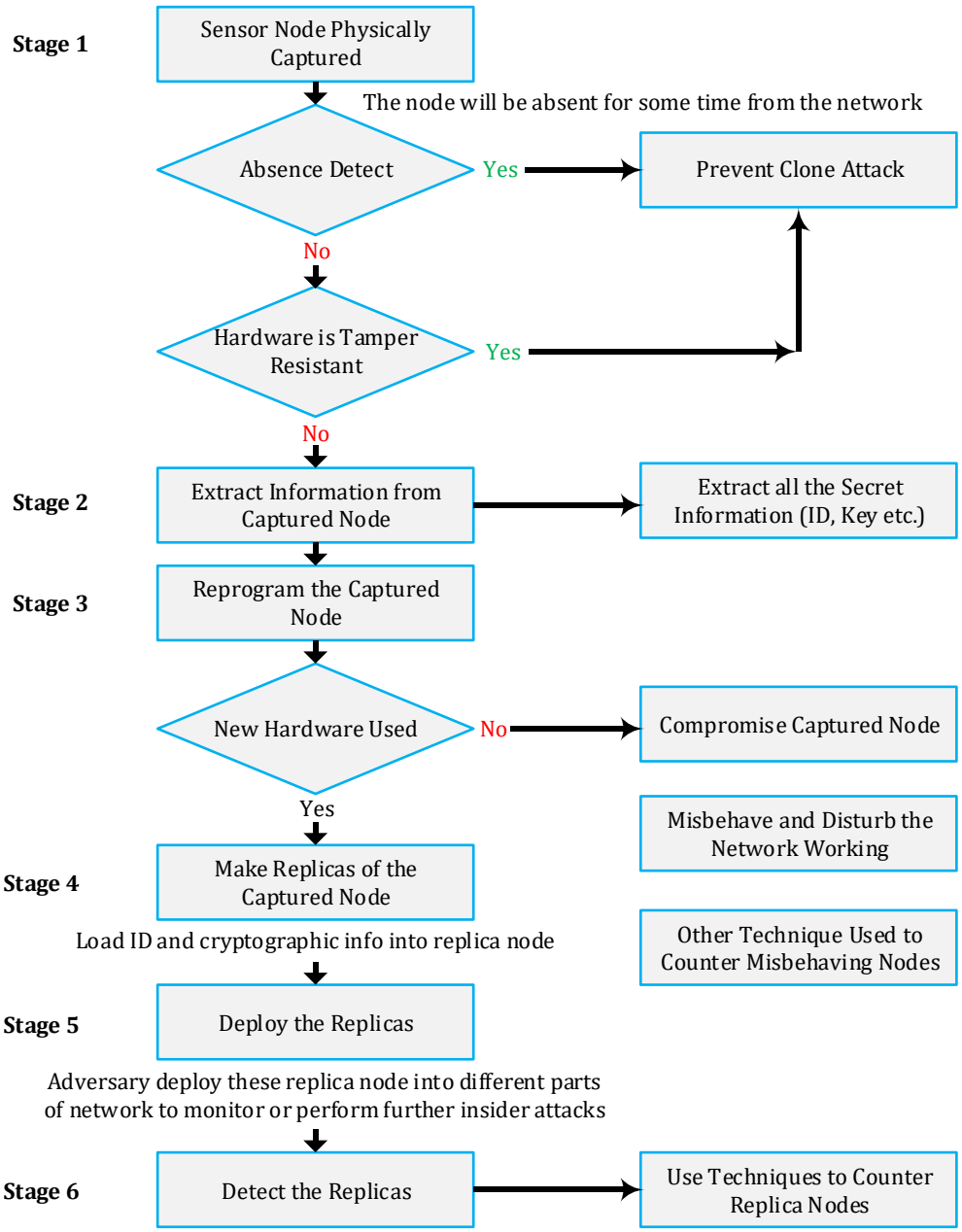


Figure 2.2: Stages of node clone/replication attack in WSNs.

2.1.3 Node Replication Attack effects on WSNs security goals

The best known security goals of all time for WSNs are: availability, authenticity, confidentiality and data integrity. In the previous chapter, it was clarified that the node replication attack is a type of attack that obviously compromises the security of the WSNs; thus, it is threatening at least one of the four WSNs security goals. Of course, this would happen if the attack was not detected from the first stage, which means that no techniques or schemes for detecting clone attacks are used to detect from the first stage. The adversary would definitely succeed in creating node replicas using legitimate IDs of existing nodes. As a result, the other nodes would never treat them later because intruders would communicate with them normally. This is particularly because these clones have all the information and cryptographic material extracted from the legitimate nodes that were captured in the first phase [41], and carry out a series of internal attacks as DoS and others of what has been mentioned earlier.

Focusing on the four security goals of WSNs:

1) Availability: Availability is very important for the WSNs' network service to continue [42]. When the adversary captures any legitimate node, he/she can be able to make it disappear from the whole network. This makes it easy to start a DoS attack or to disrupt legitimate signals. If the network is available, there is still a chance that an attack will be detected before further internal attacks are started.

2) Authenticity: Authenticity is a technique that is designed to enable the node to identify the other sensor nodes communication to make it sure they're legitimate. The problem is that in the case of node replication attacks, the adversary uses IDs for existing legitimate nodes when creating clones, which makes detection more difficult. This is how network authenticity is being affected by clone attack.

3) Confidentiality: Confidentiality is one way to ensure that only those who have a legitimate right access the data on the network. However, if a node is captured and a clone is created, they will function as if they are legitimate. The created clones thus access the data. The problem is that the data may contain business trade secrets, government data, or even private medical or

financial records, etc. The data, when retrieved from illegal bodies, then the data might get into trouble. This is how the confidentiality of data is being affected by the node replication attack.

4) Data Integrity: Data Integrity is all about making sure that the data is transmitted safely from the sender to the receiver without any changes or without being intruded by any attacker. Yet, during the node replication attack, the adversary is able to pledge in some false corrupt data and send them to the receiver; subverting them through the created clones. This is how an adversary steals the data, falsifies it and re-sends it again. The problem is that this process takes place within few minutes.

2.2 Related work

From what has been mentioned in the literature on security issues that threaten static WSNs, node replication attacks have received the most attention and research in this area. For this reason, the literature is full of designed of clone detection protocols, all of which fall into one of two categories: centralized and distributed techniques.

This section focuses primarily on identifying the characteristics of sensor nodes that increase the likelihood that they will be affected by many physical and internal attacks, and on the role of clone detection techniques in ensuring the security of WSNs mentioned in the literature. These techniques and approaches are discussed in more detail in this chapter, while outlining their advantages and disadvantages.

The main goal is therefore, to identify the challenges and research gap from the literature through SLR by critically analyze the centralized and distributed clone detection protocols.

2.2.1 Node replication attack detection in static and mobile/dynamic WSNs

Static WSNs differ significantly from mobile ones in that the sensor nodes of the static WSNs are being deployed randomly in the network with fixed positions. However, the sensor nodes of the Mobile Wireless Sensor Networks (MWSNs) move freely in the network and do not have any fixed positions. They are also able to control their movement, interact with the physical environment around them and collect the information they need. Thanks to the advances in robotics that have contributed to the development of these mobile sensors, they have autonomy and the ability to capture, compute and communicate just like the static WSNs. This even helps them to easily exchange the data / information they have received, as long as they are within the

same area on the network. However, this could be seen as the main difference between the two types of WSNs.

In addition, static WSNs use fixed flooding / routing for data distribution, unlike mobile WSNs that use dynamic routing. Furthermore, it is fairly obvious that static and mobile WSNs would have different clone detection techniques; mainly because they have different properties. To make it clear, the sensor nodes of the static WSNs have fixed locations where they are deployed, it is easier to detect node replicas or clones are present. This would be done by recognizing that a logical ID of a legitimate node is associated with more than one node in the network. However, this would be quite difficult in mobile WSNs; where the nodes do not have a fixed position and are roaming in the network [43].

2.2.2 Clone attack detection mechanisms in statics WSNs

The main problem lies in the fact that clones behave like the original legitimate authorized nodes. This is actually done using all of the authentication techniques and network communication protocols that are present in the WSNs, and even allows the clones to create paired shared keys with the legitimate nodes and to encrypt / decrypt their communications [38-40]. Thus, there are clones that cannot be detected easily in the networks which operate legitimately as the other nodes. This process does not cost the adversary; where he/she uses the available tools [37] to launch attack and creates his apparently legitimate clones of a single sensor node.

After it has been proved that the available detection protocols and techniques are not that good in dealing with the WSNs attacks; especially the node replication attack. Many researchers proposed new protocols to solve this problem, all of which could be classified as network-based or radio-based as mentioned in [44], detection techniques, as explained by [45]. The network based detection techniques could be also divided into two types, one is static WSNs and the other for the mobile ones; where each could fall into centralized or distributed detection techniques, as shown in Figure 2.3. As mentioned earlier, this thesis is focusing on the static WSNs detection techniques.

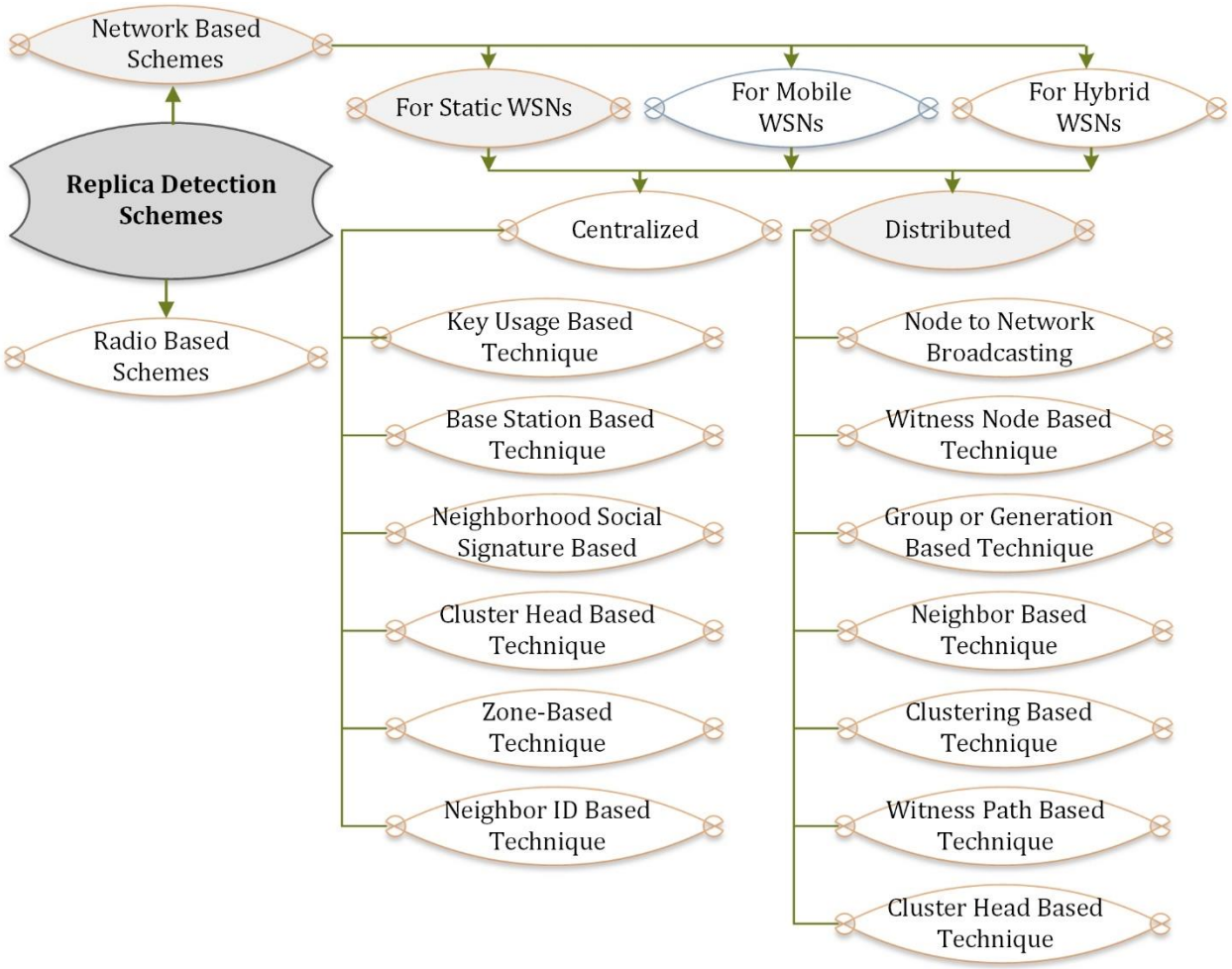


Figure 2.3: Taxonomy of clone detection schemes.

2.2.3 Centralized based detection techniques

These techniques are mainly based on a powerful Base Station (BS) for the convergence and decision making of information where the nodes send their position claims to the BS with the help of their neighbors. Then the BS would check the node IDs so that if an ID is found in more than one location, an alarm warns to announce the presence of a clone attack. These techniques have great capabilities in detecting clone attacks. However, this does not mean that the private information of the sensor is secured. Here, the attacker could do many malicious things to spy on the information transmitted between the sink node and the sensor nodes. As a result, the network may still be at risk. Another problem with the network is that its lifespan is rapidly decreasing due to the fact that those nodes closer to the sink node lose their energy faster. The most crucial

problem is that this type of method for clone detection may be cause of single point BS of failure (the attacker can take interest into make replica of the BS).

In general, all centralized static detection techniques for static WSNs can fall into one of six categories i.e. key usage based technique, based station based techniques, neighborhood social signature based techniques, cluster head based technique, zone-based technique, neighbor ID based technique, shown in Figure 2.4.

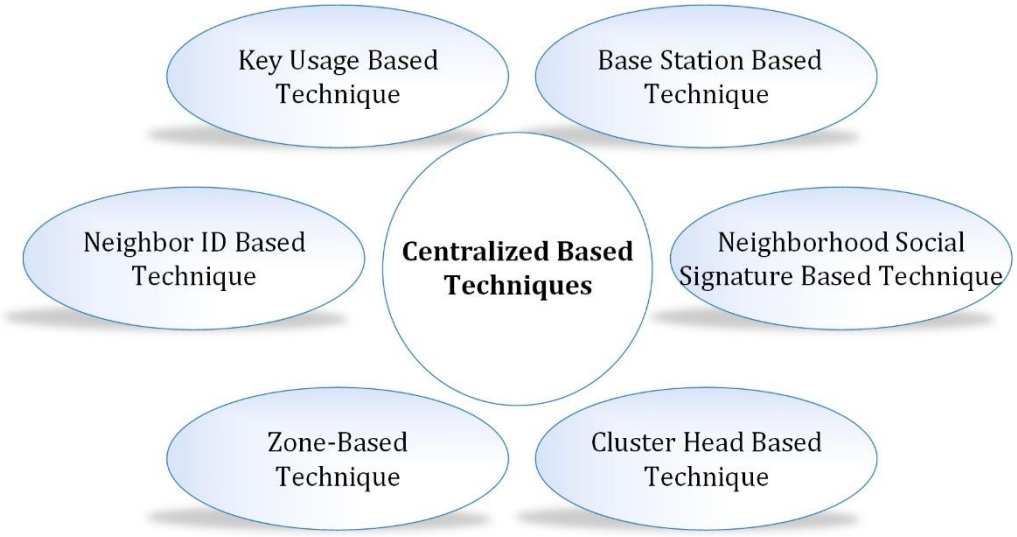


Figure 2.4: Centralized based detection techniques

1) Key usage based technique

The random pairwise key pre-distribution method are introduced in [2] for centralized based detection in which the keys are working according to a certain pattern of random keys pre-distribution method, where exceeding the threshold of keys consider to be clone. In this method the Bloom filter for counting the keys are used to calculate the key usage statistics. In the network every node executes the Bloom filter procedure for keys which are uses for communication with neighboring nodes. It attaches random number to Bloom filter and encode the result with the public key of base station, then the encoded data forwarded to base station. After receiving the Bloom filters, base station decrypt it for discards the duplicates by counting the number of time each key used in the network. Once count keys, the value above from the threshold are considered to be clone. When find out the clone base station then create Bloom

filter from of clone keys, encode the list with secret key and send the filter to all nodes in the network using gossip protocol. Next every node decrypt the Bloom filter received from the base station, discard the clone keys from own keys list and terminate the connections from the clone node keys. Afterwards every node constructing Bloom filter of k symmetric keys (random selection of keys from the large pool) for detection and send to the BS. Once BS station receive then count the total number of times of each key used, in the result if keys are often used are considers to be clone and then revoke from the network.

This method is seeming like effective, if only the size of key pre-distributed to every node is lesser and more clones present in the network, hence will find poor detection accuracy. Besides, it is supposed in the method that the links are possibly equivalent between nodes. Although in WSNs, the sensor nodes can communicate only with a limited number of neighbors practically, in a fixed wireless communication range. One more flaw of this method has ignored to guarantee that the joining clones report their keys reliably to BS.

2) Based station based techniques

These techniques mainly rely on powerful Base Station (BS) for information convergence and decision making; where the nodes with the help of their neighbors send their position claims to the BS. Then, the BS would check the node IDs so that if one ID is found in more than one location, an alarm alerts to announce the presence of a clone attack.

SET operations technique was proposed by [32] and it depends on creating sub-trees and sub-sets for each sub-tree that all forward their reports and information to the base station for clone detection in the network. The idea is all about randomly dividing the nodes on the sub-sets, and then the sub-set leader would collect the nodes' information. Afterwards, the sub-set leader would send this information to the root of the sub-tree it belongs to. The sub-tree would later compare all the data it got to detect if there is any clone node. If there is no clone in the sub-tree, all the sub-trees would send their reports to the base station to detect if there are any clones in any of them. In other words, it would detect if there are more than one node with the same ID in more than one sub-tree. The major problem facing this solution is that it needs high costs for the message authentication codes, in terms of computation and calculation. This operation is also too complex to be used and could have a reversed effect and instead of detecting the clones, the adversary might use the same protocol to revoke the legitimate nodes.

The Randomized, Efficient and Distributed detection (RED) protocol has been proposed in [46, 47]. During the first step, the nodes along with the base station will all share a random value, called rand. Then, the second step comes which is the detection phase, so that the base station would detect if there were any clones in the network. This phase would be done when the nodes send their location claims to their neighbors who would then send these claims with a probability "p" to a set of pseudo-randomly selected network locations. Thus, here it is like having more than one detection phase and that the replicated nodes would be detected in each phase. Although this solution seems great and efficient, it still has some weaknesses; especially that it has a very deterministic selection characteristic for witness nodes, in each round of the protocol. The main problem lies in the fact that the distributing random seed of the RED might not always be available for this job to be done. Besides, the adversary/attacker might be able to either compromise or even avoid the witness nodes, which would end up succeeding in creating the clones.

Compressed sensing Identification (CSI-1) technique was proposed by [48] for static WSNs. This proposed solution works by broadcasting a fixed sensed data (α) by the nodes to their one hop neighbors. These data are to be sent to the sensor nodes, and then they would be forwarded and aggregated along the aggregation tree, whose root is the base station, through the compressed sensing based data gathering techniques. These aggregated data and results would afterwards be received by the BS and the data of the sensor network would also be recovered. Finally, the node which is found with a sensory reading greater than the (α) will be the clone, which means that there is a number sent more than once and all legitimate nodes could only send their numbers once.

Efficient Centralized Approach for clone detection is proposed by [49]. This solution/approach is relying on the BS' digital signatures, and that the BS would look for a new node and give it the task of sharing pair-wise keys with its neighbors. This approach would detect clone attacks after finishing 4 steps/phases, which are neighborhood discovery, agreement demand, agreement grant and finally PK computing, network authentication and new code authentication. This authentication would be eventually guaranteed through the new node's key, as well as the BS digital signature. Then, all the nodes would have to send an encrypted notification (i.e. send their data/identity) to the BS, so that clones could be detected easily.

However, this approach has many of the weakness points that other centralized approaches suffer from. Besides, this approach faces a huge problem in being inflexible and incapable of system expansion.

Later Compressed Sensing Clone Identification (CSI-2) approach has been proposed [6] and it mainly relies on the state-of-the-art signal processing technology. This approach is based on the idea of countering the node replication attack in static WSNs by aggregating the compressed readings of sensors. The main unique advantages of this approach are having a low communication overhead, as well as evenly distributing the network traffic over sensor nodes and extend/expand the network operation lifetime. This is done after destroying the clone's sparse property in sensor nodes. This approach is composed of three phases, which are System Initialization, Structure Configuration and Clone Detection. The second and third steps take place the same way it happens in CSI-1. Yet, the only difference lays in the first one or more specifically in the fact that the CS recovery is not performed on the aggregated measurement vector as in CSI-1, but here the BS itself looks for the matching aggregated measurement values in the Val-column of lookup table L through a Binary Search. If (Vec,Val) is found as a result of this search, then, there is a clone vector in this network. This is how the clone attack is being detected using CSI-2.

Two algorithms Position Verification Method (PVM) and Message of Verification and Passing (MVP) have been proposed in [50]. According to the study the MVP is important for clone attack detection for analyzing the trustworthiness of nodes. Then, these nodes should be verifying through the PVM, but not for each and every single node. This approach paid attention to the value of time and programmed the PVM application to check for authentication before communication to save more time. PVM algorithm is mainly used for checking the data and identity/information of each node, as well as the transmission of data throughout the network. This would all be checked using the INODEINFO table of the BS. The most important point is to make sure that each node is authorized by the BS and if not, then this node would not be able to communicate with any other node in the network. Then, after registration, the PVM algorithm would compare the data/IDs/locations/timestamps it got from before communication with the current ones to check if there are any clones; where the PVM algorithm results show only the trusted nodes; to make sure that the data is transmitted in a secured way; where the node could

only communicate with others through passing its verification message. Afterwards, if any node has been suspected, the authentication message would be checked through comparing the PVM with the MVP algorithm. Finally, the clones would be stopped and another path would be selected for them after detection.

3) Neighbor social signature based technique

Fingerprint and Verification is a neighborhood social signature based technique for detecting clone attacks (i.e. nodes replicas) in static WSNs, which was proposed by [51]. The idea revolves around computing fingerprints for sensor nodes through the neighborhood information. This would be done using super-imposed s-disjunction code [52]. Then, whenever a node is sending a message to any other node in the network, or transmitting data, it would include its fingerprint in that message. This fingerprint would indicate its location (in which neighborhood it belongs) as well as its ID. If a node's fingerprint is found in a place different to the neighborhood it is located it, then it is being detected as a clone. However, this technique is not perfect, yet weak and has many limitations besides those that centralized solutions already have. This technique is weak to the extent that any clever attacker could compute a fingerprint for a clone by himself which is consistent with the neighborhood it belongs to. This would make the detection part more difficult. Besides, a replica/clone could even not communicate with the BS from the first place; so that it could not be detected.

4) Cluster head based techniques

Area-Based Clustering Detection (ABCD) has been proposed by [53], and was compared to the Line Selected Multicast (LSM) technique. This is to accomplish the desire of achieving the highest successful rates possible, yet with the lowest communication overheads as was the main aim of creating the new tiny cheap sensor nodes. This technique is working with choosing/selecting a Central Node, which is the node having the maximum number of neighbors related to it. As shown in the sub-title, this technique has clusters; thus the network itself would then be divided into sub-areas, each with a selected witness node in each. These sub-areas would be located at equal angles around the central node. Afterwards, each witness-node would receive the location-claims of the nodes in its sub-area, and if any witness-node found what is called a location conflict (i.e. a problem with the location claim of any node that it is suspected to be a clone), the witness-node would send a conflict notification to all the nodes in the network. If not,

then each witness-node would send the location-claims of its sub-area to the central node to check and compare them on a network level. The problem is that this approach has a high communication overhead in the inter sub-area of clone detection. And also although the sub-areas could be scaled through dividing their angles, this approach could easily fail. Regarding calculating the communication overhead, it is first important to note that the highest storage overhead is found in central and witness nodes; where the location-claim is sent first to the witness then to the central nodes.

Another solution for clone detection is cluster head selection approach which is a hierarchal algorithm based approach that is proposed in [54]. This solution would rely on a Bloom Filter Mechanism and would detect the network reactions towards the node replica attacks. Besides, this cluster head selection would be done using the LNCA (i.e. local negotiating clustering algorithm) protocol [55]. The process itself goes as follows; where node replicas are being detected through exchanging the member nodes' IDs of each cluster head among other clusters in the network, using the Bloom Filter, in three consecutive steps/phases. The first phase is mostly concerned with the distribution of the needed material for Bloom Filter consumptions and cryptographic operation throughout the network among the sensor nodes. Then, the cluster head election takes place in the second step. Afterwards, during the final step, the construction and verification of the Bloom filter are being performed by the cluster heads.

Secure Cluster Election (SCE) and Secure Efficient Centralized (SEC) Approaches clone detection approaches have been proposed by [56]; where the SCE is used before the SEC approach. In the SCE, what happens is that the cluster heads do their best to prevent ant node replication attacks through a Paillier method. Then, each cluster head would communicate with the base station. Afterwards, the role of the SEC approach arrives; where it would check if there are any node replicas in the network through a central Base Station authority. For clone detection, this approach uses three different scenarios; where the first is simply based on the idea of working with static WSNs and that obviously the node are being static and located in fixed position in the network. Thus, these fixed positions are to be stored in their neighbor node that would automatically observe the nodes' locations, IDs, as well as the distances between them and the neighbor node while communicating with each other. Then, the neighbor node would compare these data with each other and if it found them illogical or not associated with each

other, then it informs the cluster head that there is a suspect for being a clone which is the role of the SEC algorithm to be used by the BS for clone attack detection. The second scenario is that the node has been already replicated and place in another cluster; thus the cluster head that the original legitimate node belongs to would not be able to detect the replication. Yet, the neighbor node would find it, send it to the cluster head to check its information, and then the BS would be informed and immediately remove this replica from the network. The third and last scenario is that there is a random node selected at time (T) to send its information (ID, location, number, etc) to the Base Station through its cluster head. Then, the BS should use the SEC approach to identify if there are any replicas in the network this way.

5) Zone-based technique

This has been proposed by [57] and it depends on dividing up the network into several "zones", each with a "leader"; where all the zone leaders share with each other their membership lists (i.e. which zone has who) for them to detect node replication attacks. The process takes place in two steps; where in the beginning each node should be registered through the ZONE REGD that the leaders send to the network nodes after being deployed. Then, each node would send back a ZONE JOIN message to its nearest zone leader. This phase/step is called the "Zone Registration" phase. The second step/phase is called "Replication Detection" and its protocol takes place on two levels; where the first one is the Intra-Zone detection level and it is carried out by the zone leader to check the node ID in its membership list if any node sent its "ZONE JOIN" message later than the rest and if this ID already exists, then it means that it is duplicated/replicated. At that point, the zone leader would send a zone revoke message for node replication attack and remove that clone from the network. If the node

ID is not on the membership list of the zone that it sent the ZONE JOIN message to, then the zone leader would ask for the help of other zone leaders to check if this node Id exists in their zones to determine whether the node would be added to or removed from the network. Thus, it should be noticed that this approach does not require nodes to send any location-claims, which means that locations are somehow independent here. This is considered a good solution for the storage overhead for it to be much reduced. However, it still has another weakness point which is that the zone leader itself might be replicated by an adversary; so that the wireless network would definitely be compromised.

6) Neighbor ID based technique

In this research the Extended, Randomized, Efficient and Distributed (X-RED) method has been proposed by [58], and it is based on RED. Thus, it would be logical to state that the design of X-RED is very close to that of RED. However, there is still a difference; where the witness node is selected dynamically for clone detection by a randomized hash function. It also performs its task through fixed time intervals. This approach consists of two steps; where first of all, the node should sign its ID and location-claim for them to be sent to the farthest neighbor in a direction that is selected randomly. Then, when the neighbors receive this information from the nodes, they check the authentication of each and every message received and verify the signature of the node sent this message. Besides, this information is to be extracted and checked by the witness node every time a node sends a message. Also, all neighbor nodes' information is stored within a diameter. What happens is that if a message is sent with a certain Id and location for the first time, then the node stores this message; however, if another node within the same diameter sent another message or is found by any means within the same diameter, then it is detected as being a clone/replica; especially if its claim is not coherent with another one stored in memory. If there are no detected clones, then a new node in the same direction and close to the boundary of a circle would be chosen to start the same procedure from the beginning again to detect clones, and so on.

Table 2.1: Comparison of centralized based detection techniques.

Year & Reference	Scheme	Algorithm	Communication Cost	Memory Cost	Advantages	Drawbacks
2007 [2]	Key usage based	Brooks et al. Scheme	O(n log(n))	--		Higher rate of false positive & negative, how to guarantee that malicious nodes are reliably report their secrets to BS is not addressed
2007 [32]		SET	O(n)	O(d)	Location independent, Lesser memory necessity	Single point of failure, Costly
2007-2011 [46, 47]		RED	O(√n)	O(d √n)	Lower memory overhead, Higher detection probability,	Need a trusted entity (Trusted third party is needed),

					uniform distribution of witnesses due to Pseudorandom selection of the witness nodes	Deterministic
2012 [48]	Base station based	CSI-1	O(n log(n))	O(log(n))	Highest probability rate for detecting replica nodes	Communication and storage overhead are high
2015 [49]		Tayeb Kenaza et al.	Trans, to the BS*NS+ response of the BS*NS	2 keys (IK U and P)	Achieved high clone detection rates	This approach suffers from the lack of scalability and also shares other common drawbacks of centralized solutions
2016 [6]		CSI-2	O(n)	O(1)	Lowest Communication and storage overhead	
2017 [50]		PVM- MVP	$O(N^2)$	O(N)		Throughput of the network is higher due to which the network lifetime is decreasing, time consuming and cost effective
2008 [51]	Neighborhood Social Signature based	K. Xing et al. Scheme	C. (1 + ratio)	O(d) + min(M wlog2 M)	Low computation overhead	Can't grip a sophisticated clone which can smartly work out by itself, a fingerprint dependable by its neighborhood for detection and protection
2012 [53]	Cluster Head Based	ABCD	O(n log(n))	O(n)	Probability of detection is high	Single point of failure, High communication overhead, decrease network lifetime
2013 [54]		LNCA and Bloom Filter	$O(t^2)$	O(t)	Communication overhead is comparatively low	Detection probability is very low
2018 [56]		SCE and SEC	O(n log(n))	O(n)		
2013 [57]	Zone-Based	ZBNRD	$O(N.\sqrt{n}Z)$ +	O(d)/O(nZ)	Dynamic	Dynamic

			O(Nz.√ N)		detection of replicas	detection of replicas, Deterministic
2014 [58]	Neighbor ID based	X-RED	O(n log(n))	O(n)	Detection probability is higher, decreased the memory overhead	Large traffic overhead

2.2.4 Distributed based detection techniques

The main difference is that the clone detection process is performed by all network nodes, which means that no central authorization node is assigned to do the job. This also means that the nodes that are located at relatively distant positions in the network are also involved in this process. Focusing on the static WSNs, there are seven different types for detection techniques i.e. node to network broadcasting, witness node based techniques, group or generation based techniques, neighbor based techniques, clustering based techniques, witness path based techniques and cluster head based techniques shown in Figure 2.5.

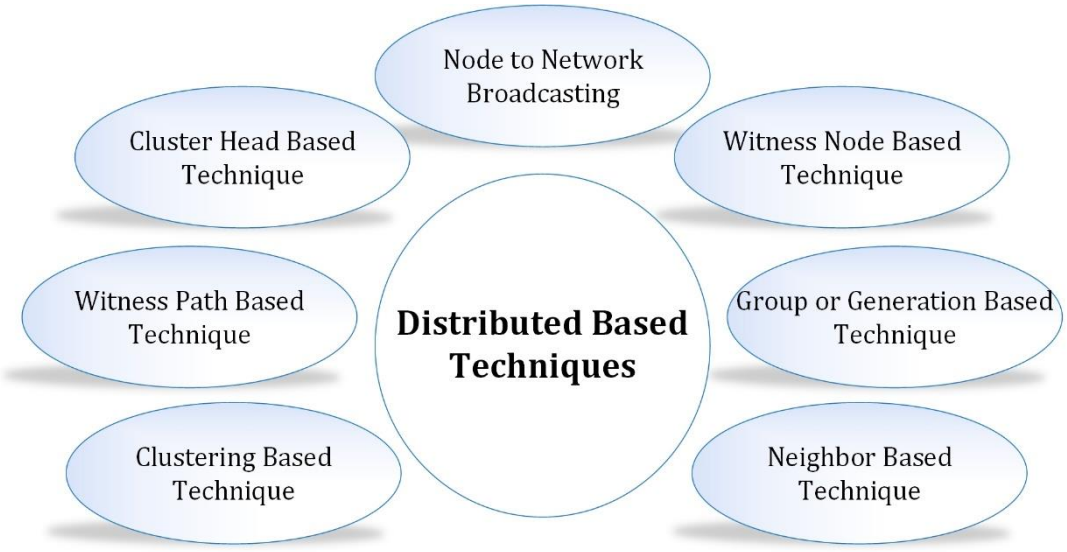


Figure 2.5: Distributed based detection technique

1) Node to network broadcasting (N2NB)

It was first proposed in [59], and it is considered one of the simplest decentralized clone detection attacks' approaches for static WSNs. Not only that, but also this approach is known of being 100% effective and able to detect every single node in the network by making sure that the broadcast reaches all the nodes. Here, nodes do not send their location-claims to their neighbors

who would then send them to a central authority or a Bs to check for clones. However, each and every single node would claim its own location by itself. This would be done with the help of putting-in an authenticated broadcast in the network. Then, each node would store all the information of its neighbor nodes, with a storage cost O(d). Afterwards, each node would wait until it receives at any time a conflicting claim and would immediately send revocation messages against the suspects (i.e. felonious nodes) for them to be isolated form the network.

2) Witness node based techniques

As the previous approach, the Deterministic Multicast (DM) one too has been proposed in [59] as the first proposed Claimer-Reporter-witness Based Distributed Framework for clone detection in WSNs. It should be stated that this solution is to an appealed and highly used one and also not the best one with the best expected results; however, it was proposed mainly to reduce the communication cost, which was the main aim of developing the new tiny and cheap sensor networks as was mentioned in the beginning. The solution to the communication cost problem proposed here is to send the location-claims of the nodes not to all the nodes, but instead to be sent to a limited determined and critically chosen set of nodes to perform as witness nodes. What happens is that the claimer node would send its location-claim to the witness nodes through its neighbors that would act as reporters, using a function. Then, the witness nodes would keep an eye on the location claims and information they receive from the nodes in the network to detect if there are any conflicting claims that announce the presence of clones in the network. The problem here is that the adversary could easily be able to get the information of the function and get over the witness node and let his clones escape from detection. Yet, this approach found a solution for this problem too through increases the effort and the time that the adversary should spend to get over the witness node. This would be done through deploying (g) instances of a function in the network; where each claimer would be mapped to different witnesses, not only one.

Randomized Multicast (RM) and Line-selected Multicast (LSM) probabilistic algorithms have been also proposed by [59]. These are full-fledged witness node based techniques for clone detection attacks in static WSNs, based on the claimer-reporter-witness based framework, like the DM. However, they could be dealt with as a development/improvement of the DM. First, in the RM, teach claimer node sends/broadcasts its location to its neighbors after signing it. Then,

those neighbors would verify this location and would become reporters themselves with probability (p). Afterwards, these reporters would randomly choose (✓) locations in the network to send the authenticated location claim to those nodes near-by the random destinations which are the chosen witness nodes for detecting node replicas. Finally, if any witness node received two conflicting location claims, it means that there are clones in the network. This would be discovered when the reporters of those replicated nodes choose the same random destinations. The probability of this scenario to happen is very high. And of course, this witness node would then inform the whole network with the presence of a clone to be discredited and revoked. The major problem with RM is that its communication cost is really high; as it needs each and every node to send its claim and the rest to be reporters, and so on. In LSM, the same procedures go on; where the claimer sends its location to its neighbors who would then become reporters and send this location claim to the randomly selected destinations (i.e. witness nodes), with probability (p) after being locally checked in the same neighborhood; to get checked for signatures and authenticity in the network. The difference here from the RM is that on the road of the location-claim from the reporter to the witness nodes, all the other nodes in the network that meet-up with this claim would store it to act as further witnesses in the network. This would be done through drawing a line across the network itself. Thus, if there is a clone, there would be one node at the intersection point of the two paths of the conflicting location claim. This is how the clone is detected here using the LSM. It is important to note that this algorithm has been proposed to reduce the communication costs in the network and to also increase the probability detection of clones. In other words, it is to propose a better solution and an improved one rather than the RM.

As an attempt to improve the detection probability of clones in WSNs, [60, 61] proposed two protocols called Single Deterministic Cell (SDC) and Parallel Multiple Probabilistic Cells (P-MPC) that are even better than LSM. The detection would be carried out through dividing-up the network into smaller cells; where each cell would have nodes and those nodes would all be dealt with as being possible witness-nodes. Thus, these two protocols are about creating geographic grids within the network. First, the SDC works in the following pattern; where every node ID is meant to be in a certain cell, and then the claimer-node would send its location claim to a specific cell. This location-claim would be, afterwards, received by any node within that single determined destination (i.e. cell) and that node would be the witness node and would save that

location claim. That witness node would then send this location-claim to that unique cell through a geographic hash function (Not in paper 77). The P-MPC is quite different; where the node's location-claim is sent to more than one cell, unlike the SDC. However, the other steps go the same as they are in the SDC. Now, talking about the critical points that pop-up whenever these two protocols enter a discussion or being thought of to be used, the choice of the best size (s) for the cells is the most important thing. This is because choosing a too large cell size (s) would lead to increasing the communication cost too much which is needed to be avoided. On the other hand, choosing a too small cell size (s) could make it easier for the adversary to capture and compromise all the nodes in the network; thus would result in facilitating the attack instead of detecting it. Moreover, there is a problem with the SDC where to reach the target of reducing the "broadcast overhead", it actually needs to make sure that the first node that receives a node's location-claim becomes the "witness node" for the SDC to get the very first copy of the node's location-claim. Otherwise, the SDC would find a huge difficulty in differentiating between the legitimate node and the replicated one.

Memory Efficient Multitask (B-MEM, BC-MEM, C-MEM, CC-MEM), these four "witness-node" based node replication attacks detection protocols have been proposed by [62]. Taking them one by one, the first proposed protocol is B-MEM (i.e. memory efficient multitask with bloom filters). This one is about compressing the information stored in the WSNs with the Bloom Filters, as well as randomly sending the location-claim of a certain node to other locations through the node's neighbors who would probably become "witness-nodes", along with that one node which is so close to the randomly selected location that the claim is sent to. These witness nodes would be able later to forward the location-claims of any node to any other random location within the network. Those witness nodes would be the ones responsible for detecting the presence of any clones within the network. Thus, they would receive a "conflicting claim" whenever there is a clone in the network and would then broadcast it throughout the entire network to revoke this clone attack. The second protocol is the BC-MEM, which is "memory efficient multitask" with both bloom filters and cell forwarding techniques. This protocol is working with first dividing up the "deployment area" into "virtual cells"; where each and every cell has an anchor node assigned to all the nodes in the network; in order to receive the location-claim of each and every node. This location-claim would be transferred through the anchor nodes then until it reached the final cell. Not only that, but not all anchor nodes are the same; where

those of the first and last cells are considered witnesses, while those of the intermediate cells are dealt with as watchers. Now with the third protocol, which is C-MEM, or the memory efficient multitask with cross forwarding technique, and it is quite obvious from its name that here there is a cross point (a random one of course) that is chosen in the network to later receive all the location-claims of the nodes in the network. Those location claims would be sent along horizontal and vertical lines till they reach the cross point that assigns the closed node to it to become the witness node, while those along the horizontal and vertical lines would become watchers. Finally, the fourth and last protocol here is the CC-MEM, which is the memory efficient multitask combining both the cell and cross forwarding techniques. This is considered the ultimate solution for the overhead cost reduction while clone detection, as well as the center and crossover problems.

Randomized Directed Exploration (RDE) has been proposed in [63], and this protocol is a witness-node based one for clone attack detection in WSNs. The idea is not new; where there would be neighbors who are randomly selected to receive location-claim messages and information from the nodes with their other neighbors'' information. This information would then be sent in a specific direction/destination which the immediate nodes have to follow to detect clones/node replicas in the network. This is mainly because these nodes already stored the location-claims and can receive conflicting claims announcing the presence of some clones in the network. And this is simply how the detection process goes, and could be successfully carried out in dense networks.

A Resilient and Efficient Replication Attack Detection Scheme for WSNs is a distributed, deterministic, witness-node based technique that has proposed by [64] to help detecting node replication attacks in WSNs. This protocol takes place in three different phases/steps, which are: Initialization, Witness node discovery and Node Revocation. Starting with phase 1, which is the Initialization phase, it starts before the deployment of sensor nodes in WSNs. This phase aims to verify the sensor nodes before being deployed in the network and this would be done through creating a verification point (vp), which is the target location coordinate, by the Base Station (BS) for each and every node ID using the geographic hash function (F). This vp could even be determined before deployment with experience by the network operator itself to a certain extent. In the second phase, which is the witness node discovery, the replicated nodes in the network

that have the same IDs as those legitimate nodes but located in different locations in the network would be detected through the location claim messages. Finally, in the third phase which is the node revocation, the witness nodes would send revocation request messages to the BS after suspecting the presence of some clones in the network. Then, the BS would check these requests to make sure if they had been encrypted by the witness nodes using the pair-wise key or not. If yes, then the BS would send a list of the replicated nodes with the reporter node itself through the network. If not, then the BS would consider the reporter node as a compromised one by the adversary and that the adversary was the one who sent the revocation message to the BS not the witness node.

Random-Walk-Based Approach (RAWL) and Table-Assisted-Random-Walk Approach (TARWL), these two protocols have been proposed in [65]; where the second one which is TRAWL is actually based in the first one which is RAWL but they are not the same. Focusing on the first, RAWL is closer to the proposed protocol in this paper and for clone attack detection, it mainly depends on the component of Randomness; where for each and every claimer-node, there are several random walks taken in the network. The procedures are not brand new; however, they actually look quite similar to a lot of the protocols mentioned earlier. Things start when the claimer-node sends its location-claim to its neighboring nodes who would then randomly select some other nodes in the network to probabilistically send these location-claims to. Each of these nodes would afterwards take random (t) walks/steps, and the passing-nodes would be later chosen to become witness-nodes to save the location-claims. The detection occurs when any of these witness nodes receives a different location-claim for the same node with the same ID. The TRAWL protocol is close to the RAWL but it aims to reduce the memory overhead, using trace table.

Note-Based Randomized and Distributed Protocol (NRDP). It might be the first time during this paper to mention a protocol based on a note. But here, it is really different; where [66] have proposed a new mechanism for detecting clone attacks in the WSN. In this mechanism, there would be a note that has the group/subset of the claimer neighbors. It would also have a selection for a reporter node from the neighboring nodes. This “reporter node” would then start receiving some requests from the claimer nodes to get a signature note. Afterwards, the claimer node would receive his signature node and would go get verification for it, then sends its location-

claim to the reporter node that would send that claim to the already selected witness nodes. And as usual, if there was any conflicting claim found, then the node of the claim is considered a clone (i.e. replicated node). It is obvious how this mechanism spends too much on selecting the reporter nodes, which is the base of this scheme. It is mainly because the whole scheme relies on the member list of the subset of the claimer neighbors for detecting node replicas in the network. This is not something guaranteed to a large extent and carries some risk within it for sure; however, things could go smoothly and easily through verifying the trust worthiness of any claimer node after checking/listening to its broadcast message. It is somehow the most trusted method to how things work in this scheme.

Distributed Hash Table Protocol (DHT) and Randomly Directed Exploration Protocol, these two proposed protocols for detecting replicated nodes in WSNs have been designed and explained in [67]; where the first one which is the DHT is a decentralized protocol and it is designed for capturing the clones from the network. It is mainly based on key caching and a unique system to continuously check for the presence of these replicated nodes. However, the second one is a randomly directed exploration protocol that continues the work of the first one. What happens in the whole mechanism is that the claimer nodes would send messages in the beginning to a random set of neighbors in the network, with a maximum hop limit. These messages would then be regulating and transmitted through a probabilistic directed technique. This would guarantee a more random, efficient and resilient performance, even in communication, in facing node replication attacks by any adversary even the smartest of them. What would help more in that is the maintenance of a property line through the network that is created as a result of this technique. Moreover, this protocol has found a solution for the communication overhead reduction, which is the usage of a border determination mechanism. Regarding clone detection, it is mainly the job of the intermediate nodes in the network; where they would detect and check the claiming messages that were sent by the claimer nodes in the beginning. Then, the role of the DHT appears; where in each round, there would be an action message with a random seed that needs to be sent by the initiator, before the round starts. Afterwards, there would be an observer assigned for each neighbor node; mainly to create the claiming message for the node by itself. This observer would be called the examinee and is quite independent and able to send the claiming message with a claiming probability for communication overhead reduction; especially in high node-degree networks. Moreover, the

hash concentration value of the random seed of the DHT, as well as the examinee ID would both be the message key that determines the DHT route/destination in the network. This means that there would be a claiming message to be sent to a destination node determined by the message key of the DHT. This message would go and check for any cloned nodes in the network by checking the ID-Location pair of the nodes. Thus, the determined destination node would be considered as an inspector, just like many intermediate nodes in the network; to strongly face any smart adversary who wish to infuse some cloned nodes

Here, there are two witness node based proposed verification protocols, which are the GDL (i.e. Global Deterministic Linear Propagation) and the RMC (i.e. randomized Parallel Multiple Cells Linear Propagation). Those have been proposed [68] for clone detection in WSNs. First, the GDL scheme is responsible for propagating and saving the information of the nodes (IDs and location-claims) along the horizontal and vertical directions in the network; where whenever the horizontal and the vertical lines intersect, this means that there is either more than one node in the same location or the same node ID in more than one location. On the other hand, regarding the RMC scheme which is based on the combination of both linear and localized multicasts, there is the component of randomness in the process of clone attack detection. What happens is that there would be selected limited regions in the WSN from which a random selection for some witness nodes takes place. Each and every place/region is known as a cell, which is a line-selected one. Then, in each cell, there would be some witness nodes along some x-axes and y-axes for clone detection. Afterwards, there would be a Birthday paradox to help mapping the location claims of the nodes till reaching the arbitrary verified cells for clone detection. If there are any nodes in the network, then a collision would be found with a very high probability. Regarding the weakness points of each scheme, the GDL, because of having very deterministic verification procedures, it is not that flexible whenever there is a smart clone attack and it is not strong/robust enough to detect it. Thus, the RMC is considered a stronger and better extension of the GDL to face the smart attacks. The best thing in the RMC is the randomness component; where it really increases the compromise-resilience and security abilities. However, the major problem is with the Birthday Paradox; where the collision does not always appear after mapping the location-claims whenever there is a clone in the network. This high probability is not always here.

In [69, 70] the authors have proposed techniques for the detection of node replication attack in WSNs. In [69] the authors have proposed RWND (Random Walk with Network Division) which is a distributed solution to detect node replication attack in static WSNs. In this technique, a claimer-reporter-witness framework is followed and for the selection of witness nodes, a random walk is employed within each area. This technique not only detects clones efficiently but also ensures high security of witness nodes with moderate communication and memory overheads due to the division of the network into levels and areas. The authors in [70] have extended their RWND technique by employing a new kind of constrained random walk and called it SSRWND (Single Stage Memory Random Walk with Network Division). SSRWND is an amalgamation of single stage memory random walk and network division which aims to decrease the communication and memory costs while keeping the detection probability higher. The proposed techniques are not suitable for the small network as when the network is divided into more areas the number of sensors nodes in each area will decrease. therefor the attacker can easily comprise the small number of nodes and neutralize the detection effects of the proposed algorithms. Similarly, for a particular network, it is very hard to find the optimal number of areas and the total number of random walks to be instated in each area for possible interaction. Obviously, the number of areas and more random walk will ensure high security but ultimately it will result in high communication, memory and energy cost.

Energy Efficient Ring based Clone Detection (ERCD) technique has been proposed by [71]. It is a ring based technique mainly because there is a ring area in the network that has the location information of both the sensors and the witness nodes of the network that were randomly selected. This property is to ensure the legitimacy of these nodes and sensors and to check if there are any replicated nodes. It is also to somehow distribute the witness nodes within the whole network; thus balancing energy consumption. This technique is considered an energy efficient one; as here the information is being transmitted between the sink and witness nodes, and the process takes place in only two steps. The first step is the witness node selection one and it is for the witnesses to securely receive some private information from the network sensors. The second step is about legitimacy verification; where there is a verification message sent to the witness node along with the private information it gets from the sensors. Then, the witness nodes would send those verification messages to the witness headers for further verification. Afterwards, the witness header would start comparing the messages it received from the witness

nodes with the already stored records for clone detection. If there is any duplicated message, then there is a clone.

Clone attack detection via Pair Access Witness Selection Technique (PAWS) has been designed and proposed by [72]. This is for more effective clone attack detection. This technique replies on having common nodes as pair-mates (ex: witness node). The idea is that each and every single node has its location-claim and each one broadcasts a signed one. Besides, every two nodes are paired as mates. Moreover, each node has neighbors; thus, each node always has to have an updated list of its neighbor and then compares and retrieves the common nodes on this pair mate's list. Then, the common witness nodes would be chosen and would receive the location claims of the nodes in the network. Afterwards, in search for clones in the network, the node ID and location would be checked by each witness node and with the help of the XOR operation that has been already performed in the beginning while creating the pair mates. If any witness node received a conflict claim (i.e. same node ID but in different locations), then this would alert the presence of clone attacks in the network, and these nodes would be blocked. Finally, the revocation process would take place, which is the Frequent Level Key change.

RE-GSASA has been proposed in [73] for clone attack detection in WSNs and it is quite different from the rest of the proposed protocols. This is because it is based on a Residual Energy as well as GSA-based Simulating Annealing. Here the technique is to aggregate data in WSNs using the maximum residual energy possible with the sensor nodes. This would be like the Cluster Head (CH). Since this technique is a location-based one, then it is important to understand how the location is important here. And it is important in the sense that the probability rate would increase by giving the location-based clone attack on the cluster node. Looking in depth at how the clones are being detected, the network nodes would all be divided up into two main groups first, which are the: witness and claimer groups. This division would make it easier and more efficient detection. It is also important to note that the nodes in the network would actually be part of both groups periodically. What happens then is that the nodes in the witness node group always send some messages of acknowledgment with time stamp. These messages would later be received by all the neighboring claimer nodes along with the single range of the witness nodes in the network. Afterwards, the nodes in the claimer group would reply to the acknowledgement messages sent by the witness nodes by forwarding response

messages that basically consist of: node ID and the pilot. This would take place in a time interval that is approximately less than the network channel coherence time, even if the message delivery was delayed. These response messages, with the help of the node ID and the pilot, would help in detecting the suspected nodes to be then sent to a new group called the suspected group. After being transferred to that group, the nodes that are suspected to be clones have to be verified along with the attack itself, which would only be done by forwarding the nodes in the suspect group to the group of the claimer nodes. The nodes in both groups would then be compared to each other and if more than a node found with the same ID, but different and random sequence, then the nodes would be considered under a clone attack. Since speed is the key in detecting clones, the witness nodes are usually being selected using the GSA-based Simulating Annealing technique that is able to get the fastest optimum algorithm, yet with the highest convergence. The GSA aims to find out the best nodes to be put in the witness nodes group to better send the request messages to the claimer nodes in the network, with a timestamp. The aim of the Gravitational Constant is actually to make the search more accurate until a point is reached where no further search is needed. Finally, the detection process starts and that is best to be done through observing the neighboring nodes' behavior for getting out the suspects and putting them in their suspects' group to be isolated from the rest of the network. Then, they would be revoked if two incoherent reply messages are sent to the witness nodes in the network. Thus, this is how the detection process takes place. The main role of the GSA after selecting the best witness nodes in the network is also to both: enhance the packet delivery ratio and decrease the packet drop. What also makes this protocol different from the rest is the procedure itself. Here, when one witness node (n1) receives whatever information and records is needs from the neighboring nodes (n) in the network, it starts doing its job by checking this data, until it finds more than one node with the same ID. This would be the kind of node information it would store. Then, this would be further checked by a broadcasted random sequence that is actually send by all witness nodes to their neighboring nodes. The witness nodes when they send these sequences and requests are actually waiting for a reply from the nodes to start the detection process, and they somehow differ from one witness node to another.

3) Group or generation-based techniques

A group/generation-based protocol that has been proposed by [74, 75]. Here, the focus is on the deployment of nodes in the network in the first place to be restricted through the usage of a

defined group-based deployment model, as well as a symmetric polynomial pair-wise key establishment model. The idea is quite simple; where yes as the rest of the protocols, the nodes would be deployed, yet this time successively in groups/generation; where each of the nodes knows quite well how much is there in each group/generation. Then, after deployment, each one of these nodes would be able to create pair wise keys with its neighbors in the network. The detection of clones occurs because clones are actually those nods in the old generation that was unable to create pair wise keys with their neighbors. Thus, this is how they are being detected and revoked.

The group deployment knowledge-based protocol has three schemes for clone attack detections in WSNs that have been proposed by [76]. These three are basically the: basic scheme/approach, the location-claim approach and the multi-group one. Beginning with the Basic Scheme, here the nodes do not really communicate with nodes in other groups. Thus, here nodes are looking for the trusted nodes of their group only, and accept nothing (nor messages nor information) from any other group. Although it seems a lot of restricted scheme, it is good for reducing computation, communication and memory overheads. However, there is still a problem; where the deployment points of the nodes are too far from the group which makes it difficult for them to communicate. This leads to a very poor network connection that would definitely be unable to be deployed in highly resilient appliances. Then, the second scheme, which is the location-claim, has been proposed as a solution for the previous approach. It is more flexible; where t allows the nodes to communicate with those in other groups as long as they are able to prove that they are legitimate nodes and not replicas. Yet, this requires the network clones to be already detected for their locations to be known. What is really impressive with that approach is that it has even less communication, computation and memory overheads than the previous one, although it has a better clone detection capability. Yet, this does not mean that this scheme is a 100% flawless one. It still has some risks and problems and the greatest one here is that there might be some flooding fake claims that might lead to DoS (was mentioned earlier in this paper). The third and last approach is the multi-group one and it is clear from its name what it shall be doing. It basically tries to make it quite harder for an adversary to compromise nodes in the network by making each node to send its location-claim to nodes in more than one group beside its original one, so that whenever the adversary decides to attack the network, he would have to compromise many groups, not just one, this would be quite difficult and would consume a lot of

time and effort and would kind of reduce these kinds of attacks. However, the problem with this scheme is that it really has high communication overhead.

Distributed Detection Resilient to Compromised Nodes in WSNs is a group-based protocol for clone attack detection in WSNs that has been proposed by [77]. It is quite different from the previously explained techniques; where this protocol in specific does not rely on the presence of any trusted entities in the network to prevent the node replication attack. Here, each and every node in the network is required to engage in the detection process. This would be done through using a one-time seed for the detection of each clone. Besides, each node would be assigned both a turn number and a seed for its turn. A new stage would then begin when it is time for the node to start the detection process; where it has to send its seed and ID with its signature first to be verified by other nodes in the network. If the verification succeeds, then the nodes would execute the detection process which would be done after dividing up the nodes in the network into groups, which would be responsible for the detection process, for more flexibility to the replicated and compromised nodes in the network (i.e. clones). Thus, the group would be able to carry out its work during its turn and detect the existing clone only after the survival of at least one of its nodes. While assessing this protocol, it would be fair enough that it makes it harder for the adversary to carry out his/her clone attack in the network. This is mainly because of the fact that the adversary would have to first spend some time knowing which group has the next turn to perform the detection process; to compromise its first node to know more about the location of the group of the witness node that would carry out the next detection process.

4) Neighbor-based techniques

Neighbor-Based Clone Detection Scheme (NBDS) has been proposed by [78] for detecting node replication attacks in WSNs. Its idea is the simplest ever and it has the logic of what real people would do in their communities and how they would deal with their new communities with their new neighbors if they got some. What happens is that the new node would tell his neighbors more about his identity and the place he came from and his neighbors would go and ask the new node's ex-community and neighbors to make sure that it is a real legitimate node. If it is a real one, then it would be verified. If not, then it would be considered a node replica that threatens the network. This is typically how this protocol detects clone attacks in WSNs.

5) Clustering based techniques

An Intrusion Detection Mechanism in Clustered WSNs has been design by [79] for detecting node replication attacks in the clustered WSNs. It is mainly based on a clustering scheme, which is known as NI-LEACH. To understand how this protocol works, it is important to know that this too depends on a random selection of witness nodes from the network as in most of the other mentioned schemes. But here, they do have a very important role, which is doing the intensive computation and energy consuming jobs that are required to be finished for clone detection. Although the witness nodes would soon, after the intensive work they would do, have no more energy left, this scheme is quite helpful and has many advantages. It is best and main advantage is that it is flexible enough to choose the most suitable encoder functions to be used depending on the requirements desired for this job to be accomplished.

6) Witness path based

A Low Storage Clone Detection Protocol (LSCD) used in Cyber-Physical Systems has been proposed and explained in [80] for clone detection in WSNs. Here, there are some witness nodes in the network that create route paths as circles; where at the center of each circle there is a sink. This detail is important; as the whole detection process (of any detection route) occurs in a centrifugal direction. Besides, the length of the witness path is definitely more than that of the distance between two detection routes in the network.

7) Cluster head based techniques

Location and Trust-Based Node Replica Detection Distributed Method (LTBRD) for WSNs for clone detection in WSNs has been explained in [81] This algorithm is considered to be more efficient because of the identification method it follows. Here, the clone attack would be detected through finding out any location mismatch as well as any problems with the trust values of any node, which are known from the node ID and information the network stores. What is brilliant with this algorithm is that it could work with or without aggregation as if it is not that crucial. In this approach, there would some nodes being treated as end devices and others that would act as neighbor nodes. To understand how this technique really works, first it should be mentioned that there is a connection between both sets of nodes; where the end devices would sense and send the physical parameter to the neighbors. Then, the rest of the nodes in the network would quickly start knowing their closest neighbors after adapting themselves to the communication range they are located in. Afterwards, the nodes that are considered as the end devices would be aggregated

while the network is being processing. Here, the trusted values appear and whenever a node passes this value, it would soon be able to send its location. Up until here, there is nothing unique about this approach. The source of uniqueness lies within the detection process of the clone attacks itself; where the problem is that any replicated node might take the ID of any legitimate node within the network and just falsify any location. And if there is any location conflict found, then the node is considered a clone. At this point, the work of the LTBRD starts; where the calculation of the locations of the nodes would only be done after each node forwards its location claim to the witness nodes in the network. This is the only way that the location of any node could be authenticated and stored for the clones to be detected later on. The way this protocol works is somehow considered a weakness, which resulted in the addition of another technique to the original protocol called TRIMAP that depends on dealing with the network as if it were a pie. This pie would then be divided-up into triangles; where each triangle would take care of only three nodes. If there is a clone in the network so that there are two different location found with the same node ID, then there would be two different mid-points available for the same node ID. This is how the replicated node is being detected here in that protocol after being updated.

Physical Layer Reputation-Based Technique for Clone Detection (PRCD) protocol has been designed mainly for fixing the demerit of the Proximity Service (i.e. Pro-Se) and it was explained in details in [82] The detection process here takes place in three different steps; where the first is the reputation assessment, the second is the holistic detection and the third is the leak detection stage. The first step is the one which detects the clones in the network and this is mainly because both the legitimate nodes and the clones exist in the same cluster. This is actually the common scenario that most of the attackers follow. Yet, if the adversary is a bit smarter than the rest, then there is still the second, as well as the third steps. They depend on the idea that the Pro-Se is much stronger than the usual security services. Thus, this protocol is strong enough, as claimed, to detect nearly all kinds of clones. For the second step, the holistic detection, it is able to detect clones even if the clone and the legitimate node are in two different clusters. Then, the lead detection would be capable of detecting the attack when the cluster-head itself is the captured one. This step can also detect other unusual sorts of clone attacks. The procedure of detection itself takes place in three successive stages/phases. The first one is the initiation phase, which is clear from its name that it is performed at the very beginning. Then, there is the data transmission phase that follows. The third and final stage is the periodical clone detection phase

that results in the repetition of the second stage after it is done. Regarding the Pro-Se function that distinguishes this protocol from the rest, it is the one that triggers the third stage to start, by sending a clone detection request. Then, within this third stage, there are three procedures to follow, which are the steps mentioned in the beginning.

Table 2.2: Comparison of distributed based detection techniques

Year & Reference	Scheme	Algorithm	Communication cost	Memory Cost	Advantages	Drawbacks
2005 [59]	Node to Network Broadcasting	N2NB	O (n^2)	O(1)	Higher detection rate, Proficient than centralized method	Higher communication cost
2005 [59]	Witness Node Based	DM	O (g log √n) / d	O(g)	Up to some extent the communication overhead are reduced	Provided less security
2005 [59]		RM	O(n^2)	O(√n)	Enhanced resiliency, Hard to predict witnesses	Communication cost is higher, also has lesser detection probability
2005 [59]		LSM	O (n √n)	O(√n)	Memory efficient, less communication overhead than RM	Cross over problem and crowded center problem
2007&2010 [60, 61]		SDC/ P-MPC	O (r. √n) + O (s)	O(ω)	Low memory overhead, More efficient then LSM	Depends on trusted entity and cell size. If the size of cell is larger then occur high communication overhead, if smaller then node can be compromise easily
2009 [62]		B-MEM	O (k . n . √n)	O(tk + t' k √n)	Higher detection probability, lower memory usage	Location dependent
2009 [62]		BC-MEM	--	O(tk + t' k √n')	Higher detection probability, lesser usage of memory, Solved crowded center and cross over	Location dependent

					problems	
2009 [62]		C-MEM	--	O(t + t' √n)	---	---
2009 [62]		CC-MEM	--	O(t + t' √n')	---	---
2009 [63]		RDE	O(d . n. √n)	O(d)	Good memory overhead	If topology of the network is distorted like no way to accomplish line transmission, then the RDE turn into unsuitable
2009 [64]		Chano KIM	O(√n)	O(√n)		
2010 [65]		RAWL	O(√n log(n))	O(√n log(n))	High detection probability	Higher memory and communication cost
2010 [65]		TRAWL	O(√n log(n))	$O(1)^2$	High detection probability	Higher Communication cost
2010 [66]		NRDP	O(N.g √N)	O(g)	Simplest method for exchanging group membership information	It has an extra overhead of choosing reporter nodes
2013 [67]		DHT	O(log n √n)	O(d)	Providing efficiency in communication overhead and clone detection probability	
2014 [68]		GDL and RMC	O(√1 × √m/2)	O(√n)	To deliver higher level of compromise-resilience random verification is used	Due to its deterministic verification process it is not robust to a smart node replication attack
2015 [69]		RWND			High probability of detection	High communication, memory and energy cost
2016 [70]		SSRWND			---	---
2016 [71]		ERCD	O(h √h)	O(h)	High detection probability with random witness selection	To store witnesses it requires a small ring routing, this reduces the storage requirements of the nodes
2016 [72]		PAWS	O(3√n log(n))	$O(1)^2$	Energy	Limited

					consumption, Detection probability and resiliency	Redundancy
2018 [73]		RE-GSASA	O(n √n)	O(n)		Messages overhead are high
2007&2012 [74, 75]	Generation or Group Based	Bekara et al.	O(√n)	O(1)	This scheme is simple and incurs less communication overhead	Nodes are bound to their groups and geographic locations
2009 [76]		Basic Scheme	O(m)	O(m)	The communication, computational and memory overhead is lower	Network is poorly connected which is not meet for high robust applications
2009 [76]		Location Claim Base Scheme	O(m + d)	O(d + 2m)	Clone detection capability is high with less communication, computational and storage overhead	Flooding fake claims due to DoS risk
2009 [76]		Multi-Group Base Scheme	3 * O(m + d)	O(d + 2 * m (1 + D max))	More robust to node compromise ever since an adversary wishes to compromise several groups	Higher communication overhead
2008 [77]		Yuichi Sei	O(r)	O(r . √n)	No trusted entity More resilient	Higher communication cost, Built-in detection start time
2009 [78]	Neighbor-Based	NBDS	O(r. √n)	O(r)	Location independent	Messages overhead are high
2015 [79]	Clustering based detection	NI-LEACH	O(l(1+m'2))	O (k.e)	Balanced throughput, less delay	If multiple adversaries then it do not detect
2016 [80]	Witness Path Based	LSCD	O(n √n)	O(l/r)	The dynamic mechanism in detection route establishment ensures the high detection probability, storage overhead of nodes is relatively low	The communication cost is high

2018 [81]	Cluster Head Based	LTBRD				
2019		PRCD	O(Np)	O(1/p)	Low computing complexity, long network life time	

Chapter 3

Research Design

We conducted a Systematic Literature Search (SLR) to find the challenges and answer the questions raised in our research area. The final publication sample for the SLR consists of relevant research papers.

SLR is a protocol-based research approach conducted to shortlist and assess the most relevant studies used to answer RQs. In the case of the WSN security domain, our work is in the domain of WSNs security, SLR is a stimulating research method for data collection. For this, we followed SLR guidelines [83-85]. Details steps are as under**.**

3.1.Search string formation

The most important step in SLR is the searching and filtering process. Several studies in this regard are been following [86-89]. Following Search Filters (SFs) have been used in our study for the creation of customized Research Questions (RQs) in Chapter 1 of this thesis.

SFs 1: We derived the major search terms from the RQs. These terms are (1) Wireless Sensor Networks (WSNs) (2) clone node detection (3) centralized approach (4) distributed approach.

SFs 2: Identification of synonyms for the major terms. Wireless Sensor Networks (WSNs), clone node detection: (“clone node detection” OR “replica node detection” OR "node compromise attack"), centralized approach: (“centralized approach” OR “centralized technique” OR “centralized scheme” OR “centralized method”), distributed approach: (“distributed approach” OR “distributed technique” OR “distributed scheme” OR “distributed method”).

SFs 3: Verification of the key words in the relevant papers. (“Clone node detection”, “replica node detection”, “node, “node compromise attack”).

SFs 4: We have used Boolean operators (OR, AND) for concatenation of the search terms in the formation of the search strings. For concatenation of the synonyms, the “OR” operator has used, whereas major terms were concatenated through the “AND” operator. We used the below search string and marked them as Track.

Track: (Wireless Sensor Networks OR WSNs) AND (“clone node detection” OR “replica node detection” OR "node compromise attack") AND (“centralized approach” OR “centralized

technique" OR "centralized scheme" OR "centralized method") AND ("distributed approach" OR "distributed technique" OR "distributed scheme" OR "distributed method").

Where Track denotes search string that is intended to search the literature specific to clone node detection in the context of WSNs security.

3.2.Online search venues (digital libraries)

Based on SF4, Table 3.1 shows selected digital libraries for searching of relevant studies. Tables 3.2, and 3.4 showing details of the digital libraries. A total of 33,444 research articles have been retrieved and six papers are identified via a snowballing method. By adopting the Tollgate approach, we selected 123 papers in the first phase based on the research title and abstract via inclusion/exclusion criteria. In phase 2, we reviewed these research articles and customized further it to 37 articles.

Table 3.1: String searching

S.No	Digital Library Searched	URL	Search String (Track)
1	Science Direct	http://www.sciencedirect.com/	"Clone Node Detection" OR "Replica Node Detection" OR "Node Compromise Attack" AND "Centralized Approach" OR "Centralized Technique" OR "Centralized Scheme" OR "Centralized Method" AND "Distributed Approach" OR "Distributed Technique" OR "Distributed Scheme" OR "Distributed Method"
2	IEEE Xplore	http://ieeexplore.ieee.org/	
3	Google Scholar	https://scholar.google.com.pk	Clone Node Detection OR Replica Node Detection OR Node Compromise Attack AND Centralized Approach OR Centralized Technique OR Centralized Scheme OR Centralized Method AND Distributed Approach OR Distributed Technique OR Distributed Scheme OR Distributed Method
4	Springer Link	http://link.springer.com/	
6	ACM	https://dl.acm.org/	

Table 3.2: Track result

Search String	Digital Libraries Searched				
	Science Direct	IEEE Xplore	Google Scholar	Springer Link	ACM
Track	1,555	14,854	16,900	69	66

3.3.Inclusion/exclusion criteria

The outcome of the search string *Track* is evaluated to ensure that the selected papers are likely to meet predefined inclusion/exclusion criteria. The inclusion and exclusion criteria are designed on the basis of the formulated RQs and presented in Table 3.3.

Table 3.3: Inclusion/Exclusion criteria

Inclusion criteria
a) Research studies and articles that are relevant to formulated RQs are included in the final list. b) Research work having the keywords "Wireless Sensor Network, Clone Node Detection, Centralized Approach are included in the final list. c) The research papers/articles/books/review papers written in the English language only are included. d) The articles/papers that are published online are included in the list.
Exclusion criteria
a) Literature that does not fulfill the abovementioned criteria has been excluded. b) The researchers excluded all the duplicate papers founded. c) The papers reporting table of contents or the information regarding the proceedings of conferences, workshops are not included.

Table 3.4: Publication search details in various digital libraries

Digital Library	**Total Papers Found**	**1st Level of Inclusion/Exclusion**	**2nd Level of Inclusion/Exclusion**	**3rd Level of Inclusion/Exclusion**
Science Direct	1,555	10/1,545	0/1,555	**37/33,414**
IEEE Xplore	14,854	33/14,821	20/14,834	
Google Scholar	16,900	45/16,855	4/16,896	
Springer Link	69	21/48	7/62	
ACM	66	8/58	1/65	
Publications through snowballing	6	6/0	5/1	

technique				
Total	**33,450**	**123/33,328**	**37/33,414**	**37/33,414**

3.4.Publication quality assessment

Every paper was tested against the Quality Assessment (QA) standards shown in Table 3.5 and scored in Table 3.6. The aim of the QA was to know the quality of the finally selected research papers. We performed the QA during the data extraction phase. Every QA criterion has 3 possible values: Yes, Partial and No with marks of 1, 0.5, and 0, respectively. Detail score of the finally selected 37 papers is given in Appendix A, where R1 and R2 represent the respondent's Author 2 and 3 respectively.

For each given item QA1 to QA4 the evaluation is performed as follows:

- The research article answers to the checklist queries are assigned 1 point.
- The publications containing some of the answers to the checklist questions were assigned 0.5 points.
- If there is no answer to the checklist, queries were assigned 0 points.

Question 4 intends to seek the method(s), through which the reported challenges have been identified. If the method(s) is clearly mentioned, then it is marked as Yes = 1; otherwise marked as Partial = 0.5 or No = 0. Similar criteria are used by [90, 91].

Table 3.5: Publication quality assessment

S.No Quality Standards	Value/Score
QA1: Checked appropriateness of aims and objectives of selected research paper	Yes = 1, Partial = 0.5 No = 0
QA2: Checked whether the conclusion matches with defined objectives of the research	Yes = 1, Partial = 0.5 No = 0
QA3: Checked the core terms of the research, i.e. Clone Node Detection. Additionally, check whether these terms are defined and discussed clearly.	Yes = 1, Partial = 0.5

	No = 0
QA4: Checked the appropriateness of how clone node detection scheme have been identified in the selected papers. ***Note:*** *QA4 intends to seek the method(s), through which the reported clone detection scheme has been identified. If the method(s) is clearly mentioned, then it is marked as Yes = 1; otherwise, No = 0.*	Yes = 1, Partial = 0.5 No = 0

3.5.Data extraction process

This is an important step in SLR, in which the data is extracted from already selected research articles. The criteria adopted for extraction are purely based on RQs. The predefined rules for data extraction are paper ID, title, reference, year, research database, quality of the publication, the country where the research was performed, context, schemes, methods and also pros and cons of each technique.

Table 3.6: Detailed scores of each selected paper of SLR against the questions of quality assessment criteria.

Papers	References	Respondents (R)	QA1	QA2	QA3	QA4	Points	Mean
P1	[2]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P2	[32]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P3	[46]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P4	[47]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P5	[48]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P6	[49]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P7	[6]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P8	[50]	R1	1	1	1	1	4	4

		R2	1	1	1	1	4	
P9	[51]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P10	[53]	R1	0.5	1	0.5	1	3	3
		R2	0.5	1	1	0.5	3	
P11	[54]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P12	[57]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P13	[58]	R1	0.5	1	0.5	0.5	2.5	2.75
		R2	0.5	1	0.5	1	3	
P14	[59]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P15	[60]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P16	[61]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P17	[62]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P18	[63]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P19	[64]	R1	1	1	05	1	4	3.5
		R2	1	1	05	1	4	
P20	[65]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P21	[66]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P22	[67]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P23	[68]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P24	[69]	R1	1	1	1	1	4	4

		R2	1	1	1	1	4	
P25	[70]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P26	[71]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P27	[72]	R1	0.5	1	1	1	3.5	3.5
		R2	1	1	1	0.5	3.5	
P28	[73]	R1	0.5	1	0.5	1	3	2.75
		R2	0.5	1	1	1	3.5	
P29	[74]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P30	[75]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P31	[76]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P32	[77]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P33	[78]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P34	[79]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P35	[80]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P36	[81]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	
P37	[82]	R1	1	1	1	1	4	4
		R2	1	1	1	1	4	

3.6.Data synthesis

As per the SLR protocol, we performed a synthesis of the data extracted from already filtered research articles and created different categories of the challenges. Initially, we identified 14 challenges, but the further classification was performed and few challenges are merged. Finally,

a list of 8 challenges is identified which are discussed in Table 3.7. Out of these 8 challenges, four challenges are considered as critical with a frequency of more than 10 % while the remaining three having a frequency of less than 10 % are considered non-critical.

Table 3.7: List of the Identified challenges

S.No	Challenges	Frequency Out of 34	%
1	Communication cost	16	47.06
2	Single point of failure	12	35.29
3	Detection probability	7	20.59
4	Memory or Storage Cost	5	14.71
5	Deterministic	3	8.82
6	Redundancy	1	2.94

3.7.Results

Table 3.7, discusses the challenges along with frequency range. As discussed earlier, critical challenges are the ones whose frequency range is more than 10 %. The formula for finding frequency is the total number of challenges identified in the finally selected 37 papers, multiplied with 100 and divided by the total number of papers 37.

3.8.Challenges identified via the SLR process RQs

A final sample of 37 papers was selected and data summarized from them. A list of 8 challenges, as depicted in Table 3.7, was shortlisted through the SLR from that summarized data. Out of these 8 challenges, 4 were identified as critical according to 10 % frequency criteria.

CHALLENGE 1: Communication cost

Table 3.7 clearly indicates that “Communication Cost”, having the highest frequency of 47.06%, can be labeled as the first challenge in our findings. This is an extremely important performance benchmark for sensor network protocols, due to the fact that communication utilizes more energy than other operations in wireless sensor networks. In this research, different methods [48, 50, 53, 58-61, 65-67, 69, 70, 73, 76-78, 80] are faced with higher communication cost issue with clone node recognition method. Communication costs can be described as the average number of location claims sent and received by each node during each detection period.

CHALLENGE 2: Single point of failure

According to Table 3.7, “Single Point of Failure”, is found to be the second most quoted challenge. If centralized based approaches are only taken into account, then the percentage varies of this challenge maybe 100 %, which turns out to 35.29% if both centralized and distributed based approaches are considered. In the centralized schemes, nodes are sending their neighbor node’s information to the base station which performs as a single trusted unit. The detection of clone occurs when the same ID with different locations is sensed. After detection, the clone is quiet from the network. This makes the approach simple and straightforward. However, it is subjected to a single point of failure [32, 46, 47, 49, 53] and nodes are exposed to heavy traffic near the base station. Therefore, a central node failure point may be lost if there are multiple witness nodes.

CHALLENGE 3: Detection probability

This study concludes that “Detection Probability”, having a frequency of 20.59%, can be regarded as a third-highest challenge in the clone node detection mechanism. In such a mechanism, the probability of successful detection is the most significant performance metric. It is the main security need of any detection mechanism to detect the occurrence of attack with high probability Therefore, the basic security of any detection method is required to detect the presence of an attack with high probability, while [6, 54, 59-61, 64, 79, 81] is lacking clone nodes detection probability. The probability of detection probability is described as the total number of successful detections of replica nodes during every detection period divided by total protocol runs.

CHALLENGE 4: Memory/storage cost

The third challenge, according to Table 3.7, is “Memory or Storage Cost”, with a frequency of 14.71%, which is the fourth essential performance metric. The low-cost sensor nodes have a controlled amount of energy, which makes the methods requiring additional storage to be unfeasible. In this research study, different methods [48, 65, 69-71] are subject to higher memory costs. Memory cost can be described the total number of location claims, which are kept by every node in the network.

CHALLENGE 5: Deterministic

The study shows that “Deterministic” with a frequency of 8.82% is identified as a serious challenge in clone node detection techniques. According to literature research, the witness nodes

are the most important and fundamental element in clone detection mechanism, since they are responsible for the detection of clones in the network. In the witness node based distributed schemes [46, 47, 57, 68], the assortment of witness nodes is deterministic. In such cases, the attacker can simply capture the nodes and make copies or clones of their confidential data. The choice of witness nodes ought to be nondeterministic. If the selection mechanisms of witnesses are deterministic, an enemy can easily capture and compromise them.

CHALLENGE 6: Redundancy

Another challenge in the clone node detection technique is "Redundancy" with a frequency of 2.94%. It is the absence of redundant data which is also a crucial entity in WSNs [72]. Redundancy certifies reliable data for the purpose of decision making. The reliable data is needed in the analysis, evaluation and predicting of system behavior while bad quality data can lead to flawed results in decision making. In Wireless Sensor Network (WSNs), nodes are heavily based in an area to gather information. Sensors detect similar data and forward it to the sink. Such similar data can produce redundancy at the sink. The outcome of redundant data results in more accuracy, reliability and safety while elimination aids in energy saving, as most of the energy of the sink node, is wasted in dealing with the redundant data. However, data accuracy still needs to be well-kept even if there is an increase in network cost and/or time.

Chapter 4

Proposed Protocol

As mentioned earlier, the main purpose of this MS thesis work is to solve the node replication attack problem that sensor nodes have encountered most frequently in the WSN. This kind of attacks is considered one of the most dangerous. After introducing most of the protocols previously mentioned in the literature, this chapter introduces a new, Zone-Leader based, distributed solution to detect and solve this problem.

Before presenting the solution, this chapter first explains the basic requirements that are required to design an effective protocol, as well as examines and understands the characteristics of the network in which the protocol would work with the adversary model. This would help to better evaluate the new protocol, which is based on a distributed claimer-reporter-witness based. What is unique here is that this protocol combines the advantages of both the random walk and the network division in zone techniques. The result would be a detection of node replication attacks with a higher probability and accuracy rate and a moderate overhead.

4.1.Clone attacks building blocks

Before how to detect and prevent this kind of attacks in WSNs, how the attack is being launched should be fully understood first. Here, are two building-blocks for this attack, which are:

4.1.1 Sensor node identity

At the beginning of this work it was mentioned that the WSNs contain a large number of sensor nodes. These sensors have unique identities, yet they have no tamper resistance. This will make it easier to collect the key management scheme later, making it easier for the attacker to transfer secret data / credentials, node IDs, etc. once the node captured. It is because this scheme has nearly all the cryptographic of both the public and the private keys that contain all the data needed to control the network.

When the attack is launched, the adversary usually succeeds in creating clones with the same IDs and the cryptographic material of the legitimate nodes. This ensures that the clones have the same authentication and confidentiality as the original nodes for communication. Thus, the attack

can be overcome of the if the pair-wise keys linked with the claimer node's ID. This makes it much easier to detect the clone by the legal nodes within the network to be revoked.

4.1.2 Sensor node location

The location of sensor node is very important in static WSNs, as already mentioned, the sensor nodes are fixed in the network and cannot be moved since they deployed. Thus, each node has a unique ID and deployment position. If an ID is found in more than one location, the presence of a replication node attack on the network is reported. Of course, this is considered quite difficult in mobile WSNs where they have no fixed locations.

4.2.Distributed techniques: requirements and challenges

Although the node replication attack is dangerous, it can be detected as early as the first phase of the attack, from the start. Or the attack could be detected later and the replicated nodes would be revoked from the network.

Focusing on the first stage, according to academic research [92-95] when the adversary physically captures any sensor node, it disappears from the network; where its neighbors could easily find out its absence. Thus, the attack is being detected from the beginning and the node's neighbors would then inform the whole network with the sudden disappearance of that sensor node; announcing the probability of the launching of a node replication attack. Then, when the adversary plants-in his clone, the network would successfully detect it; preventing the attacker from taking any secret data or controlling the network. This would be done through informing the rest of the nodes in the network with the presence of a clone; so that they would not communicate with it. Despite seeming simple, this step is a bit challenging; as it is not the probability of having a clone attack in the network whenever a node is absent. There could still be some absent nodes while not having any up-coming attacks. However, there could be other reasons for this absence, as the network facing some power/hardware failures or even facing some natural disasters that led to the diminishing of some nodes.

Thus, detecting the attack from the first stage is not always an easy job and does not always succeed, which means that there are a lot of attacks that had been detected after the attack is already launched in the network and the techniques mentioned in this thesis have been designed for detecting the node replication attacks after already deploying the clone in the network. Those

techniques were either centralized or distributed ones, and the distributed ones seemed to be much better, efficient and effective; especially the claimer-reporter witness-node based detection schemes [59, 65]. However, they are not 100% perfect; as they still have some weakness that had been already explained in chapter two and still lack some of the required techniques that make them more distributed and more robust and stronger in facing this kind of attacks.

Nevertheless, the following points are the ones needed to be considered while designing any CRW based techniques for better results and to overcome some of the challenges faced by any scheme in detecting the clone attack.

4.2.1 Witness nodes selection

This is an important step in designing a witness-based technique for detecting node replication attacks. If the adversary is smart, he/she may be able to grasp which nodes are assigned as witness nodes and their routes, which would make it easier for him/her to take control of the entire network. This would result in the attack being successful and the detection scheme failing.

The adversary can find out the ID of the witness nodes if the scheme is deterministic. The adversary might even know which node would be witness in which execution round. Thus, the adversary would always succeed in duplicating the desired number of sensor nodes in the network and compromise every witness node he/she meets and will obtained all the necessary secret information

The RED protocol is one of the previously explained protocols in this the literature review, which has a real problem with the deterministic selection of the witness nodes. In every round there is a seed that is forwarded on the network before the round begins, which allows the attacker to find out which node would be the new witness node and when. For this reason, the witness nodes in the network should not be selected deterministically and the IDs should under no circumstances take the security of the detection protocol into account.

4.2.2 Witness nodes distribution

The distribution part is also a bit challenging. If there is an undistributed protocol, the attacker may be able to determine the future witness nodes on the network, as well as their locations and IDs. In addition, these types of detection protocols typically focus on the selection of witness nodes from a particular geographic location on the network, which further facilitates the

adversary's work and compromise the witness node he / she want, to control the network with all the secret and vital information it stores

This problem can be found in the LSM detection protocol, where it is not distributive and suffers from a problem known as a crowded center. This means that there is a specific location on the network where all nodes are selected, and this is usually in the middle. The result, of course, is easier collection of nodes secret credentials. Therefore, the complete even distribution of the nodes in the network is very important to ensure that all nodes can later be witnesses in a uniform and equal manner.

4.2.3 Higher security of the witness nodes

Focusing on the CRW-based techniques, the witness node is considered a very important component here which needs to be highly secured. Targeted security would therefore be achieved after a protocol was created with an imperceptible ID and location.

4.2.4 Different variants detection of the clone attack

Since this type of attack is considered one of the most dangerous attacks against a network, its detection is of great importance for the survival of the network. The problem with the node replication attack is the fact that if the adversary is smarter, he/she can capture the witness nodes that are responsible for detecting clone attacks on the network. This would definitely result in capturing many nodes in the network without being detected and getting all the needed secret information.

4.2.5 High detection probability

It is even more important and crucial to securing the witness nodes after designing the detection protocol to make sure that this protocol has very high, efficient and effective detection probability; for better detection of clone nodes in the network. Detection of one clone node on the network is even more difficult and challenging where the recognition process becomes more difficult.

.

4.2.6 Moderate overhead

Usually the costs of communication and memory of any detection scheme are calculated; to determine the performance quality of the scheme. For a detection protocol to be considered as a good one with a high performance, then it should have a pretty high detection probability for detecting clones in the network, as well as having relatively low or somehow moderate costs for computation, communication and memory overheads. Not only that, but also designing a quite simple detection protocol for clone detection is important. In RAWL scheme; has a very high detection probability but the problem in this scheme is that it has very high communication and memory costs. RAWL has considered a good scheme; especially that it is a distributed, non-deterministic one, where the witness nodes are evenly distributing in the network.

4.3.Random walk in WSNs

Wireless sensor networks (WSNs), which consist of small and modest sensor nodes, have certain characteristics e.g. negligible computing abilities, memory and severe communication restrictions with a limited life time of battery. The random walk method has been used with great interest in many areas of networking, e.g. searching and checking in distributed networks, forwarding / sending and querying, processing, data collection and security in WSN [96-101]. The widespread importance and satisfactory implementation of randomized procedures is confirmed as regular, distributed, and entertaining features of random walks, such as simplicity, location, low overhead, considerable robustness against dynamic errors or changes in the system topology. No requirement for the physical system structure and topological attributes such as the area or degree of the neighboring node, transmission area, association balance and so on. With these features, many types of random walks were used in the system to randomly determine the traffic design, reorder the procedure, and cover the area data of the information sources.

Random walking is essentially a controllable method in which the nodes of a graph G can visit in some sequential random request. Some fixed nodes are chosen randomly [102], this is how the random walk begins, and then the walk moves with every step to the neighbor of the nodes. Random walk has often been used as an effective packet forwarding strategy, but there is no real lack of moderate crossfading [103].

Different types of random processes are used to identify clones in the network. Storage and calculation costs are examined in the use of random walks to select witness nodes for clone detection in WSN. With regard to the disadvantages of simple pure random walks, experts go different ways to overcome these disadvantages of simple pure random walks and examine numerous variations of random walks. A simple stage memory random walk (SSRAND) has been suggested in [102, 104, 105]. SSRAND is a kind of walk where a node can be visit on condition that node should not be the passed previous or current node.

4.4.Assumptions

In this section, we explain the assumptions used in the proposed Hybrid Random Walk assisted Zone-Based (HRWZ) node replication detection technique with random walk in zones. However, some specific notations of scheme are presented in Table 4.1 and some general notations are used to denote the expressions for simplifications of models representations, are presented in Table 4.2.

4.4.1 Network model

Sensor nodes can be easily distributing over a wide range of sensor networks. Each node can identify its geographic area using GPS or a location algorithm [106-115]. The nodes to be stationary during the execution of a clone detection protocol.

The network is assumed to be freely time-synchronized and tends to run in both centralized and distributed manner [59, 116]. In [46, 47, 59, 65] is expected that adversary will not be able to create new sensors nodes.

The assumptions made about sensor network are follows:

(i) The nodes will be static, not tamper-proof and evenly distributed in the observation area
(ii) Communication links can be single or bidirectional
(iii) There will no centralized trusted entity
(iv) Nodes will be aware of their position
(v) Nodes will be assign with a unique ID

4.4.2 Adversary model

However, for an adversary model, it is expected that an adversary can capture the sensor nodes and then compromise them. The clone node retrieved from the original node can utilize using cryptographic data, and then added to the network. The cloned node can pursue the detection protocol and in this way they don't suppress or drop messages [59]. Something else, when an opponent tries to compromise a large number of sensor nodes, the presence of verification components or automated protocols such as SWATT [117] are expected to attract human interaction and starts clearing the network to eliminate compromise nodes. Together with these it is accepted that an adversary can only select the number of nodes to be capture. It is understood that sending messages has no effect on fall or wormhole attacks, or that certain measures, such as [7, 118, 119] can be taken to counter such attacks.

The assumptions made about an adversary is follows:

(i) Limited number of nodes will be compromised by an attacker

(ii) Once node compromised, then an adversary will have the full control over the node

(iii) Using the captured node; an adversary would be able to create as various copy as he/she wishes to deploy in the network

(iv) An attacker couldn’t create a new ID for node.

4.4.3 Symbols and notations

Table 4.1: Specific notations for different schemes

RM/LSM	
N_{ls}	The line segment numbers
λ_{ls}	The line segment average length
RAWL	
N_r	The random walk numbers
N_t	The number of random walk steps
ξ_{cd}	The size of claim (bytes)
HRWZ	
N_z	Total number of zones
N_{nz}	The total number of nodes in zone
$\lambda_{zL's}$	The average path length or distance between randomly selected Zone-Leaders

N_r	The random walk numbers
N_t	The number of random walk steps

Table 4.2: General Notations

N_n	Total number of nodes in the network
N_z	Number of zones in the network
Z_g	Randomly selected nodes in the zones
$N_{zl.N}$	Number of Zone_Leaders in the network
D	Number of neighbors of each node
loc_n	Node location information
F	Size of the location claim (bytes)
τ_{sr}	Transmission cost for sending and receiving one bit
τ_{sv}	Energy cost for signature check/verification
τ_{sg}	Energy cost to sign location claim
P_f	Location claim probability
P_s	Location claim storing probability
P_d	The replica/clone detection probability
λ_n	The average path length or distance between any two nodes in the network
$\lambda_{zLıs}$	The average path length or distance between randomly selected Zone-Leaders
ID_n	Node Identity
K_n^{pvt}	Node n private key
K_n^{pub}	Node n public key
$Sig\ \{M\}_{K_n^{pvt}}$	The signature of node 'n' on message M
$\|\|$	Concatenation symbol
$ID_n, loc_n, Sig\{H(ID_n \| loc_n)\}_{K_n^{pvt}}$	Location Claim Format

4.5.Proposed protocol

In this section, a new protocol is proposed, which is based on the Claimant Reporter witness framework and can meet the needs of CRW-based techniques with moderate communication and storage requirements. The protocol is based on random walks and includes the possibility of single stage memory random walk with network division into zones.

4.5.1 Hybrid Random Walk assisted Zone-based protocol

This section briefly present, the newly designed detection scheme Hybrid Random Walk assisted Zone-Based (HRWZ) protocol for clone attack. This scheme would work on dividing the network into zones, each with a Zone-Leader; and it is a Claimer-Reporter-Witness based framework technique. What is good about this scheme is that it is a fusion of both dividing-up the network into zones and at the same time having a single stage memory random walk. Credits for creating this scheme could go to the ZBNRD [57] and RAWL [65] schemes that this chapter is inspired by a lot.

In our scheme HRWZ, for more security to be applied, there would be an Identity-Based public key system [120, 121] that would be used to make sure of the messages authentication before being sent and received among nodes in the network, as well as making sure of the signatures to know the identity of the node sending the message. This would help the Zone-Leaders, who are being randomly selected after a random walk is being employed in each zone in the HRWZ to better detect the cloned nodes in the network.

In HRWZ, clone nodes are detecting in both locally and globally scenarios. After dividing the network into zones each zone is assigned to a list of nodes belong to it. After the Zone-Leader selection, Zone-Leaders are searching for the clone nodes within the zone. This process of clone detection placed in local detection scenarios. If clones not detected locally then all the Zone-Leaders sharing their list of nodes with each other's to detect clone globally. The detection process takes place in three different steps; where the first one is the "network configuration and zones formation phase'; the second is the "Zone-Leaders selection" and the third and last one is the "clone detection", which is simply done by the Zone-Leaders locally and globally.

The process goes as follows; where in each and every execution, a random would be selected from a zone. This node would then be required to forward its ID claim to some other randomly

selected one-hop neighboring nodes called reporter nodes (remember this scheme is based on a sort of randomness). Afterwards, those neighboring nodes would be able to forward that node's ID claim to some other random nodes in the same zone. The same process goes on until the last step of random walk in the zone is reached to a random node to become Zone-Leader. It is also important to know that although selecting the Zone-Leader is a random procedure, it is still different from iteration to another in the scheme.

4.5.2 Protocol description

This section is going to present the details of how HRWZ scheme is working.

There are three steps that need to be completed for the clone detection and revoked. These steps are the network configuration and zones formation, followed by the claim forwarding, and ended by the Zone-Leader selection; to finally reach the clone detection and revocation phase.

a. Network Configuration and Zones formation

In our HRWZ scheme, as has mentioned earlier, the network is being divided up into a number of zones; where each has a leader and each leader has a list of the nodes information exist in its zone. The number of the zones in the network has to be configured during the configuration phase according to the network size. This process is known as the network zoning process, which is considered a very important process for the increasing/prolonging of the lifetime of the WSN network. The sensor nodes are being added into several zones. In each zone a single node is being randomly selected to become a Zone-Leader; to later work on collecting the information of the other nodes in its zone for detecting clones locally. While the information would be exchanged and shared with other selected Zone-Leaders for clone node detection globally. Moreover, it is important to mention that those zones are required to stored energy. One of the most important concerns of this thesis is to find a way to prolong the network lifetime, while following the zone based mechanism.

So, the first assumption is that the network coverage area N is 160m² before it is being divided up into zones. Afterwards, the network configuration step is done with the zoning process; where each zone would have the area about $16m * 8m\ OR\ 8m * 16m$ which is 128m. Finally, the pseudo-code of this whole process of how nodes are being deployed in the network is shown in Algorithm 1.

Algorithm 1 presents the pseudo-code for the dividing the network into zones mechanism in the proposed HRWZ protocol.

Algorithm 1: Network Configuration, Zones Formation and Nodes Deployment

Initialization:

1. Before connecting BS and ∀ZL's, network coverage area N ← 0 //Base Station BS, for all ∀, Zone Leaders ZL's , Network N
2. When apply 160m² then N ← 25,600m
3. **If** N = 25,600 then
4. **For** all N **do**
5. N ← Z *// Here Z used for Zones, where area of one Z =16m * 8m* or 128m
6. Z ← n *// Here n used for nodes, where in every Z, number of n =64*
7. n ← Id_i + 1 *//Here i start from 0 up to n numbers*
8. **End**
9. **End**
10. Stop

b. Claim Forwarding

For the detection process to be completely and successfully done; a claimer node in the zone that is being randomly selected first. Claimer node would then broadcasting an authorized and signed location claim to its one-hope neighbor called reporter node with a certain format $ID_n loc_n, Sig\{H(ID_n \parallel loc_n)\}$. Here, in this equation, loc_n is referring to the node's location information, while ‖ symbolizes the "concatenation operation".

c. Zone-Leader Selection

Another important step that is totally carried out by the claimer-reporter nodes is the selection of Zone-Leader. The target here is to select the Zone-Leaders, while having the lowest communication cost and memory overhead as possible as it could be.

After receiving the claim, the neighboring node will become the reporting node of the claimer node. The reporting node then forwards the location claim to some others neighboring nodes randomly selected in the zone. This process will continue till visit the last node by a random walk steps in the zone. After visit the last node will become a Zone-Leader of the zone. The Zone-Leader in the zone is responsible for clone detection locally while in globally detection of clone nodes the Zone-Leaders connecting and sharing the list with each other's shown in Figure 4.1.

In Figure 4.1, it is clear that how all the above mentioned mechanisms take place, starting from the network configuration phase. zoning process, and the selection of the Zone-Leaders via a random walk within the zone. The selection of zones is shown by the red rectangular shape, the communication between nodes in the zone are visualized by the straight black lines, the black dotted lines shows communication between Zone-Leaders, the nodes colored yellow represent claimer nodes, silk are the reporter nodes, green are the Zone-Leaders while the clone is colored red.

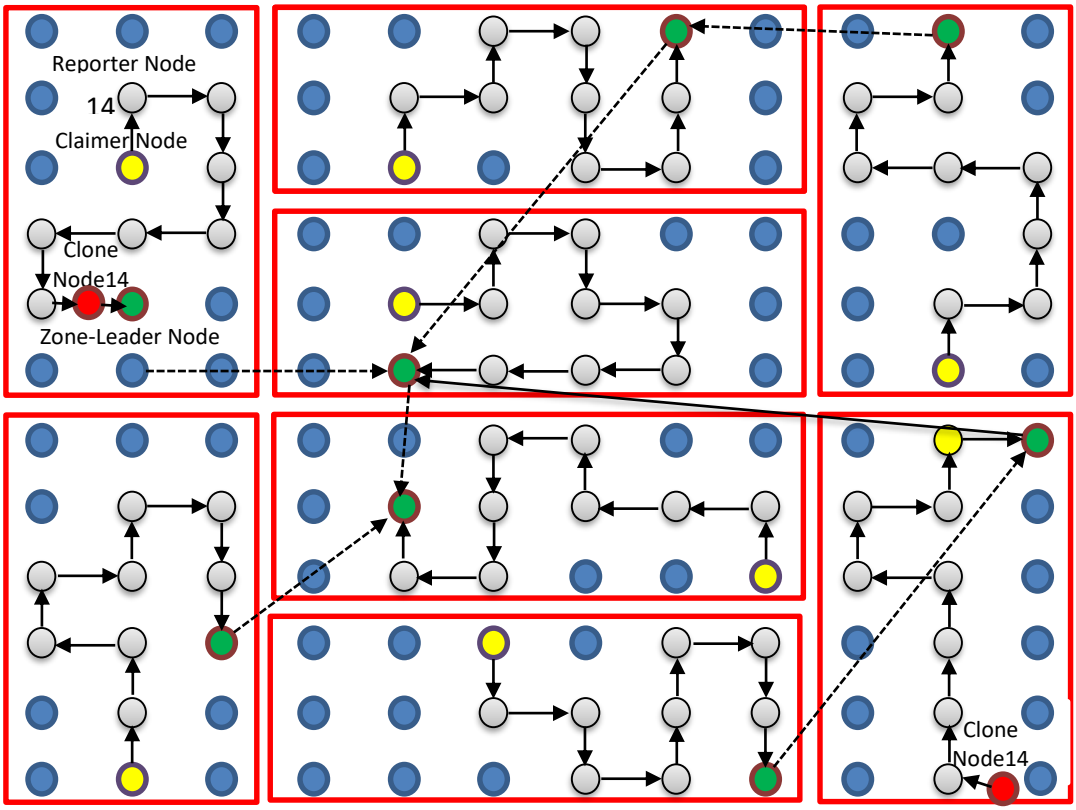


Figure 4.1: Detecting replicated nodes during random walk in zones.

As has been mentioned earlier in this chapter, the random walk scheme is a very important for the clone nodes detection. The random walk in zones is used to select the Zone-Leader having the list of all the nodes in its zone. This is for the reason to be able to verify and detect the clones nodes.

For instance, as shown in Figure 4.1, in 8 different zones the claimer nodes have been randomly selected and has start the random walk (t) with walk steps (r) in the zone till to reaches

the last nodes which are becoming Zone-Leader. After selecting the Zone-Leaders searching for the clone node in their own list, if the clone is found then the Zone-Leader is revoking the clone from its own local zone, if not found then these Zone-Leaders are sharing their own list table with each other's for finding the clones globally, if these Zone-Leader founds the conflicting location claims (i.e. two nodes with the same ID but are located at two different locations), then they will announce the existence of a clone attack to be revoked. This would be done and checked using the given formula:

Equation 4.1

$$((ID_n = ID_n)\&\&(loc_n != loc_n))$$

In Figure 4.2, there is a conflicting claim for node with ID 14 and it is detected by the Zone-Leader locally, and send a revocation message (NODE_REVOKE) for the node to be removed from the network and this message would be immediately sent to the rest of the members. Afterwards, when other Zone-Leaders, receives the lists of other zones in the network (as that of zone number 8 in Figure 4.1), it would start detecting clone nodes and verify their existence. Thus, the leader of zone 4 would verify that the ID of node number 14 exists in two different zones (i.e. conflicting claim) and would send a revocation message (NODE_REVOKE) to all the Zone-Leaders globally, as well as the nodes of its own zone, to terminate their connections with the clone node, Thus, the clones are being successfully detected and revoke.

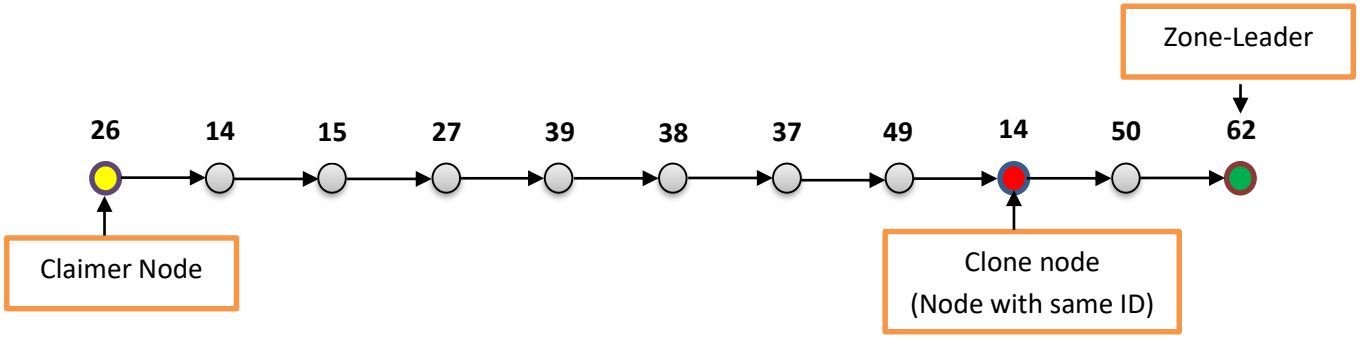


Figure 4.2: Presentation of random walk in zone 1.

The pseudo-code of all the work done by the claimer-node, reporter-node as well as the Zone-Leader in the network is presented in Algorithm 2 and clones are detected in Algorithm 3.

Algorithm 2 represents the pseudo-code of the proposed HRWZ protocol involving all the steps performed by the claimer, reporter and the Zone-Leader.

Algorithm 2: Steps performed by claimer and reporter nodes for Zone-Leader Selection

For Zone-Leader selection, in each zone a randomly selected node forwarding location claim to a random node at t++ random walk

11. **for** Z_L selection *// Here Z_L used for Zone-Leader of the zone*
12. $Z_n \leftarrow R_n$ *//Random node R_n is selected, in the zone*
13. $R_{nn} \leftarrow R_n$ (loc_n_claim) *//Randomly node R_n send the location claim loc_n_claim to a random neighbor node R_{nn}*
    $loc_n\_claim = ID_n, loc_n, loc_n\_claim, Sig_n\_claim\{H(ID_n \| loc_n)\}K_n^{pvt}$
14. **When** R_{nn} (Received_loc_n_claim) **then**
15. R_{nn} will send the R_n information to another R_{nn}
16. **do** with probability
17. t++ random walk
18. **While** L_n visit *//Here L_n used for last node in the zone*
19. $Z_L \leftarrow L_n$ *//Z_L is used for Zone-Leader*
20. **End**
21. **End**
22. **Stop**

Algorithm 3: Steps performed by Zone-Leaders for clone detection

For clone detection and revoking, Zone-Leaders verifying the nodes with same ID of different locations in the table list

23. **For** C_N detection, Identification and revoking *// Here C_N used for the clone nodes*
24. Z_L collect the data ($ID_n, loc_n\_claim, Sig_n\_claim\{H(ID_n \| loc_n)\}K_n^{pvt}$) from all nodes in the zone
25. Z_L_Verify (Nodes_Information) *// Z_L verifying the nodes information*
26. **If** $((ID_n = ID_n)$ && $(loc_n = loc_n))$
27. $AZ_{L's} \leftarrow Z_L$ (list_of_nodes_information) *//$AZ_{L's}$ all zone leaders*
28. **If** $((ID_n = ID_n)$ && $(loc_x != loc_x))$
29. Clone
30. Revoke $(ID_n = ID_n)$ && $(loc_x != loc_x))$ *//Global detection*
31. **Else**
32. No Clone
33. **Else**
34. Clone
35. Revoke $(ID_n = ID_n)$ && $(loc_x != loc_x))$ *//Local detection*
36. **End**
37. **End**
38. **Stop**

4.6.Network Division and Zone Selection

As per configuration, the whole network is divided into multiple zones and each node belongs to a specific zone. Suppose that the network is divided by the number of zones (N_z), (e.g. Z_1, Z_2, Z_3, Z_4…n). However, the division of the zones depends on many factors, in particular the selection of the Zone-Leader, the size of the zones, the cost of all communications and the size of the network. In the formation of network phase, the minimum number of zones is considered to be N_z = 8. Random Zone-Leader selection can make it difficult for the attacker to obtain information about the Zone-Leader. In the case of N_z = 8 zones, the Zone-Leaders in each zone are chosen in a completely random manner. It is therefore very difficult for the attacker to determine which Zone-Leader will be selected in the zone. As a result, maximum protection of can be achieved by dividing the network into zones.

The volume of each zone is important because the network is divided into small zones can be very easy for the adversary to recognize the Zone-Leader. It is therefore important to examine the potential size of a zone in which the network can be distributed so that the Zone-Leader's security is high. The size of each zone depends on the total number of nodes in the network. Therefore, the approximate size of each zone can be calculated by dividing the total number of nodes (N_n) in the network by the total number of zones (N_z). This can be expressed by the following formula in Equation 4.2:

Equation 4.2

$$S_z \approx \frac{N_n}{N_z}$$

Where $\{N_n > N_z \; and \; N_z = 8\}$

In the mechanisms of witness node based, the main component is the selection of the witness node. Zone-Leader in our case, and the entire process for clone detection depends on the selection of Zone-Leader randomly. The randomly selection criteria for the Zone-Leader is to protect the Zone-Leader against clever adversary. Ultimately, the primary purpose of a Zone-Leader selection randomly is to secure the network so that it is impossible or even more difficult for an intelligent adversary to compromise the Zone-Leaders.

4.7.Chapter summary

This chapter discussed ideal requirements for the distribution of clone nodes in case of static WSNs. Moreover, a new distributed clone detection technique called HRWZ (Hybrid Random Walk assisted Zone-based protocol) is been proposed. Our proposed technique follows the "claimer-reporter witness framework", which selects witnesses on a non-deterministically basis, and using parallel random walks in dynamically selected zones.

Chapter 5

Results

In this chapter we present the results and analyzes the proposed Hybrid Random Walk assisted Zone-Based (HRWZ) protocol for cloning detection in WSN. By performing the comparative study to the selected claimer-reporter witness based techniques (RM, LSM and RAWL) has been presented which validate that the proposed protocol meets to the requirements, which detect clones with the probability of more than 90% which keeps the communication and memory overheads.

5.1 Evaluation metrics

Performance is analyzed by using five evaluation metrics explained below, to assess the performance and feasibility of the proposed protocol.

5.1.1 Witness distribution

The most important requirement is the witness distribution for claimer-reporter based clone detection schemes. Research studies stated that if the sensor node memory is high or its energy is quickly used up, it dies. Therefore, to balance the storage and energy consumptions in claimer-reporter witness based techniques, the selection of witness nodes should be uniform and randomly distributed. Every node in the network should have the equal probability to be a witness node. Noted that in our case, the witness node is the Zone-Leader.

5.1.2 Probability of successful detection

The main performance metric for clone detection schemes is likely to be successfully detection of clone nodes because this is the basic security requirement of any detection scheme to be able to detect the presence of a clone with high probability. The detection probability as the total number of successful detection of clones that is detecting during each round and divided by total number of runs. By using the following formula, detection probability can be calculated.

Equation 5.1

$$Probability\ of\ Detection = \frac{(Total\ \#\ of\ successful\ Detection)}{(Total\ \#\ of\ Simulation\ Runs)} * 100$$

5.1.3 Communication cost

Communication cost is the most crucial performance metric for sensor network protocols as communication in wireless sensor networks uses more energy than other operations (137). Communication cost is defined as the average number of location claims that are sent and received by each node during each detection round.

5.1.4 Memory or storage cost

Memory cost is the another important metrics for performance. Whereas, the inexpensive sensor nodes have the limited amount of energy, therefore, techniques that require more memory are considered ineffective. Thus, the storage cost can be define is the average number of location claims that stored by every node in the network.

5.1.5 Computational or processing cost

Another metric is the processing cost used to evaluate the performance of clone detection. The processing cost can be define is the average number of the signature verifications which are performing by every node in the clone node detection process. By any node the number of signature verifications can be calculated as the number of stored messages by a node in the detection process.

5.2 Comparative study

In this section the performance of the proposed HRWZ methods with the selected claimer-reporter witness based schemes: RM, LSM and RAWL is evaluated by comparing the clone node detection probability, communication or processing cost, memory or storage cost and witnesses distribution. Our proposed scheme does not need any specialized MAC layer protocols; therefore, in the simulations for reliable transmission we implemented a simple approach RTS-CTS-DATA-ACK, and the simulations were passed out in a grid topology.

In 160 x 160 square grid network area the 1024 sensor nodes were deployed with transmission range of 5m. In case of our scheme HRWZ, the total network was divided into 8 zones while the simulation performed for 100 times run for a random walk (r) and random walk step (t), which produced the acceptable results. The grid deployment for sensor nodes is shown in Figure 5.1, while the parameters are shown in Table 5.1.

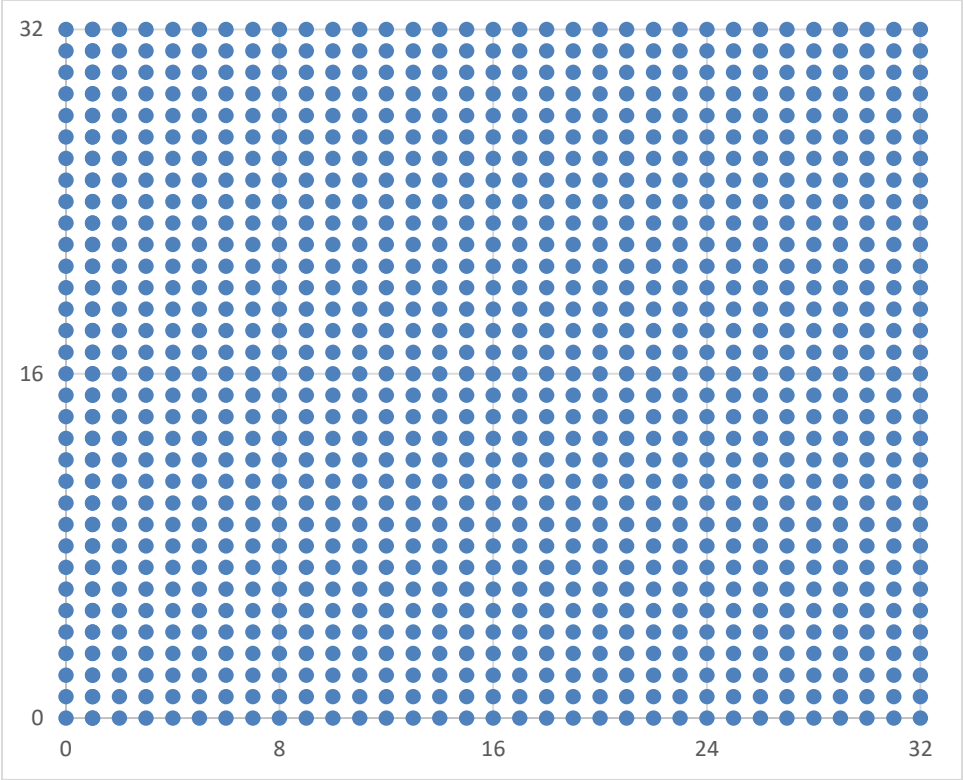


Figure 5.1: Grid deployment model for sensor nodes.

Table 5.1: Parameters for grid deployment model

Parameters	**Values**
Total area of the network	160m x 160m
Total number of nodes	1024
Network deployment type	Grid
Rang of communication	5m
Size of location claim	46 bytes
Total number of simulation runs	100

5.2.1 Witness distribution (load balance)

One of the key requirement for a witness-based technique is the distribution of witnesses, which must be fulfilled in order to guarantee a high level of security for the witnesses and a high probability of clone detection. In our scheme HRWZ, a kind of witness distribution called Zone-Leader is simulated 50 times for a single node in a zone by setting the number of random walks

(r) as 1 and the number of random walk steps (t) as 40. The Zone-Leaders are uniformly distribution in a zone illustrated in Figure 5.2, which shows that how many times a single node can be selected as Zone-Leader, where the x and y axis shows the location of Zone-Leaders.

Figure 5.2: Zone-Leader distribution of HRWZ for every zone

5.2.2 Detection probability of clone node

This section analyzes for achieving 90% detection through the needed number of walk steps for RM, LSM, RAWL, and HRWZ.

In RAWL, the probability of detection depends on the intersecting witnesses, where one intersecting witness node is enough successfully detection of clone node. As a rule, the individual intersection witness node is sufficient for the successful detection of replication nodes, while in HRWZ it depends upon Zone-Leaders and a single Zone-Leader is enough for the successfully detection of clone nodes. For the minimum communication overhead, the optimum number of random walks (r) were chosen to guarantee 90% detection probability.

Under the grid based network topology the values for detection probability (P_d) with different walk steps (t) has been chosen. For the proposed HRWZ scheme, the required number of walk steps (t) to achieve 90% detection probability, we simulated the results for 8 zones in shown in Figures 5.3. The simulation process has been run for 100 times in order to get the average value of probability detection (P_d). For RAWL, the optimal value of random walk steps (t) are 800.

Figure 5.3: Detection probabilities vs walk steps for the proposed HRWZ and RAWL

The needed number of random walk steps (t), for to achieve the 90% detection probability checked RAWL and HRWZ shown in Figure 5.3. For HRWZ when the network is divided into 8 zones, the optimal number of walk steps (t) are 100. Figure 5.3 shows that a large number of walk steps are required to obtain a 90% probability in RAWL. On the other hand, HRWZ requires a small number of walk steps in order to achieve the same detection probability as RAWL.

Table 5.2: Probability of detection values for RAWL with Random Walks (r) and Walk Steps (t)

Random Walk (r)	Random Steps (t)	Detection Probability
1	150	30%
	200	40%
	400	80%
	800	91%

Table 5.3: Probability of detection values for HRWZ (8 zones) with Random Walks (r) and Walk Steps (t)

Random Walk (r)	Random Steps (t)	Detection Probability
1	16	8%
	24	20%

	96	50%
	192	75%
	288	95%

Table 5.4: Probability of detection values for RM and LSM with Number of Witness nodes (for RM) and Number of Line-segments (for LSM)

Witness Nodes	35	40	45	50	55	60	65	70
RM	71%	80%	87%	92%	95%	97%	98%	99%
Line Segments	1	2	3	4	5	6	7	
LSM	26%	57%	77%	88%	94%	96%	98%	

Observing the overall results of the grid topology confirms that walk steps initiated in the RAWL scheme is in most of the network, which give results of high walk steps for 90% probability of detection. In comparison, of the proposed scheme HRWZ the network is divided into zones, and small number of parallel walk steps are initiated in these zones which give 90% detection probability.

5.2.3 Communication cost

Data sensing, processing, sending and receiving are the processes that consume the energy resources of a sensor node. However, the energy used to communicate in wireless sensor networks is the most important challenge.

The communication cost of RAWL involves two types of communication overheads in the clone detection procedure, the first one is forwarding the location claims from reporters to randomly selected nodes in all the network. Second one is initiating walk steps by randomly selected nodes till to the end of random walk.

For HRWZ, to reduce the communication costs than RAWL in large networks, Zone-Leaders selection method can be used. This method also involves two types of communication overheads in the clone detection procedure, the first one is forwarding the location claims by reporters to randomly selected Zone-Leaders in randomly selected zones which minimizes the cost of

random walk steps for every reporter. Second is the average path length between any two randomly selected nodes that exist anywhere in the whole network (in RAWL) is greater than the path length between any two randomly selected Zone-Leaders of adjacent zones. So selecting Zone-Leaders of adjacent zones will result in reduced average path length that consequently reduces the overall communication cost, which is finally resulting in reduced communication cost of our scheme which is the main motive of this mechanism

In RM the cost of communication involves the cost of sending the location claim from reporters to randomly selected nodes in the network. While in LSM the cost of communication involves the cost of forwarding the location claim from reporter to randomly selected nodes in the forwarding paths of each line segment.

The communication overheads of the selected CRW based schemes are illustrated in the Figures 5.4 in grid based network topology if the network is divided into 8 zones.

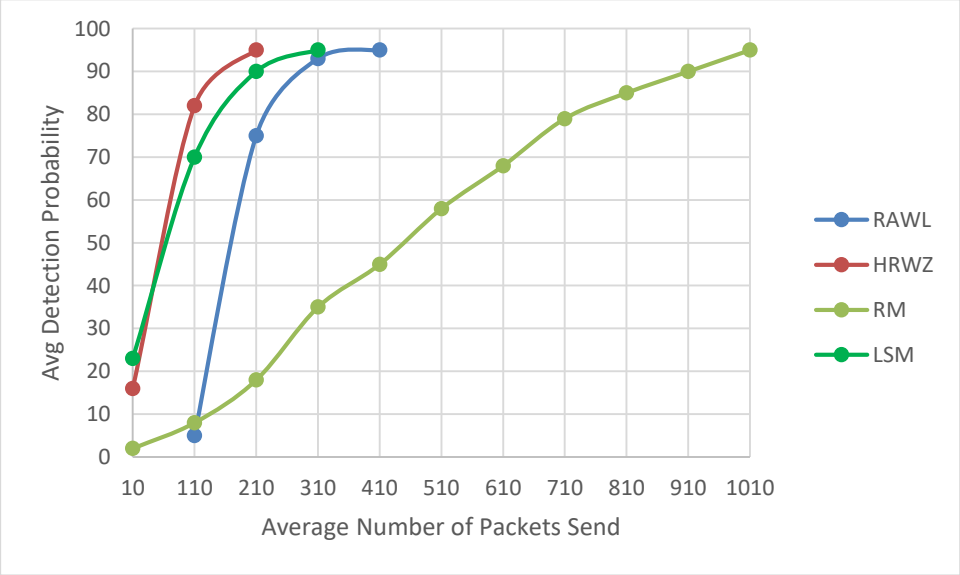


Figure 5.4: Communication overheads of the chosen schemes with HRWZ

The communication cost of RM, LSM, RAWL and HRWZ are shown in Figure 5.4, if for HRWZ the network divided into 8 zones. The results show that RM consuming the highest communication costs compared to LSM, RAWL and HRWZ. This is because the RM selecting large number of witness nodes for achieving higher detection probability, which bearing the

overhead of sending the location claims from reporter nodes to the witnesses. For achieving 90% clone detection, the RAWL needed the double communication overhead if compared to LSM, because RAWL should initiate long walk steps which cause increasing in walk steps which as a result increasing the communication overhead. As mentioned in chapter 2 that crowded center problem in LSM has been overcome by RAWL and hence trade in increasing the communication overhead for stronger security. From the results it is confirmed that if the network is divided in zones, HRWZ requires the low communication cost compared to RAWL and RM, while comparing to LSM, HRWZ has less communication overhead compared to LSM. This is due to the parallel random walk and random walk steps were initiated in these zones with much less number of walk steps for achieving the same detection probability and stronger security.

5.2.4 Memory or storage cost

Storage costs are also an important factor in the WSNs because WSNs are resource constrained networks with limited battery of sensor resources. RAWL storage costs include the cost of storing the location claim in witness nodes which are randomly selected by random walk, while the storing the location claim in HRWZ depends on the randomly selected Zone-Leaders. The RM storage costs include the cost of storing the location claim in randomly selected witness nodes across the network. LSM storage costs include the cost of storing the location claim in witness node on forwarding paths to each line segment.

The memory overheads of the selected CRW based schemes are illustrated in the Figures 5.5 in grid based network topology if for HRWZ the network divided into 8 zones.

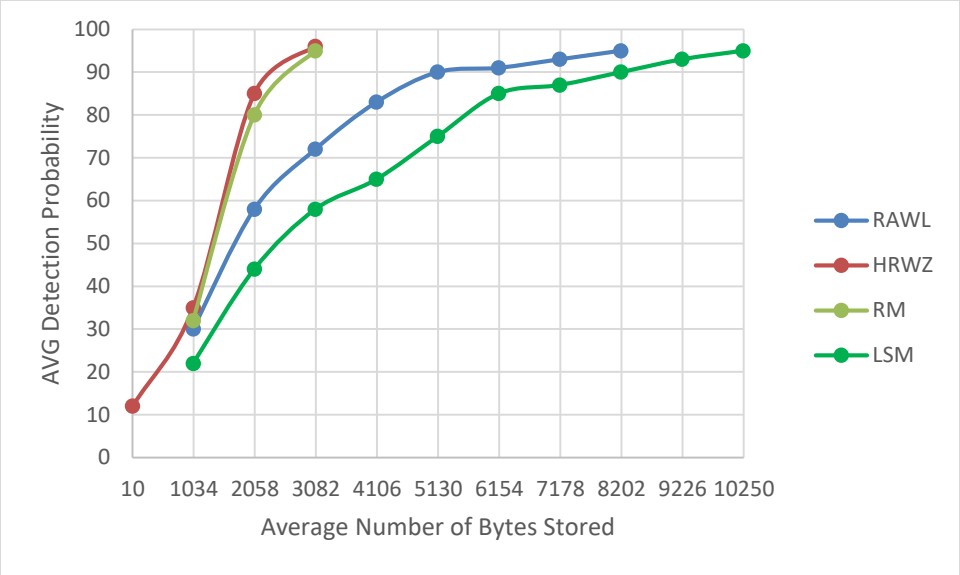


Figure 5.5: Memory overheads of the chosen schemes with HRWZ.

The memory cost of the RM, LSM, RAWL and HRWZ are illustrated in Figure 5.5 when the network divided into 8 zones. Results indicates that LSM requires highest memory overhead to achieve 90% detection probability compared to all selected schemes due to huge number of witness nodes selection which is storing the location claim alongside the forwarding path. HRWZ needs less memory overhead compared to RAWL and LSM but little good than RM because a small fraction of witness nodes is storing the location claim in RM.

5.2.5 Computational or processing cost

The cost of computation for RAWL comprises the cost of verification of the location claim by witness nodes while in HRWZ the cost of computation comprises the verification of the location claim by the Zone-Leaders which can be selected through the random walk and random walk steps. The cost of computation for RM includes the verification of the location claim by the randomly selected witness nodes in the whole network and the cost of computation for LSM includes the verification of the location claim by the witness nodes on each line segment forwarding paths.

The Computational overheads of the selected CRW based schemes are illustrated in the Figures 5.6 in grid based network topology if for HRWZ the network divided into 8 zones.

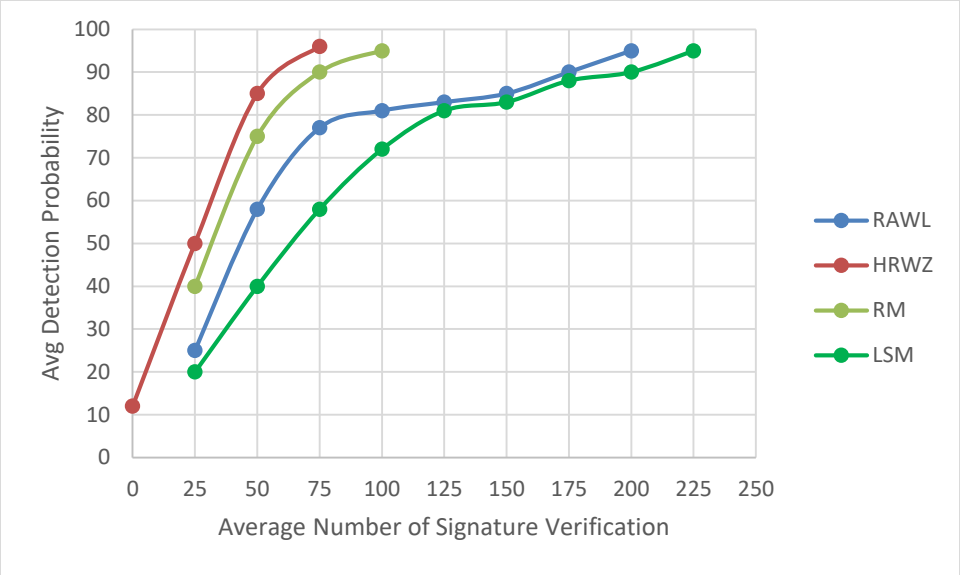


Figure 5.6: Computational overheads of the chosen schemes with HRWZ.

The cost of computation for RM, LSM, RAWL, and HRWZ are illustrated in Figure 5.6 if for HRWZ the network divided into 8 zones. Results indicates that LSM needs highest computational costs compared to others selected schemes because of the large number of witness nodes selection which verifying the signature and store the location claim. The cost of computation for RAWL needs less computational cots than LSM but higher than HRWZ and RM because of the large number of witness nodes selection in the whole network in order to initiate random walk with long walk steps. In contrast, RM requires less computational cost compared to LSM and RAWL schemes because of the selection of small number of witnesses which store location claim. Results indicated that HRWZ needs less computational costs than RAWL, RM and LSM.

5.3 Chapter summary

This chapter presented the results and discussions for a comparative study. Comparative study was performed using extensive simulations in grid type of deployment topology. In grid based deployment topology the sensor nodes are deployed in a uniform manner and the nodes are placed in a row and column fashion. The simulations result shows the comparisons of communication, memory and computation overheads of HRWZ (for 5 and 7 randomly zones selections) with the existing CRW based techniques, i.e. RM, LSM and RAWL. The results

verify that HRWZ consume less communication memory and computation overheads than the existing CRW based techniques.

Chapter 6

Conclusion and Future Directions

This chapter summarizes research contributions, limitations and future directions.

This thesis focused on a harmful clone or node replication attack in static WSNs. In this kind of attack an adversary physically capturing a legitimate node in the network, then creating clone with the same ID of legitimate node. Afterwards, the adversary would create several clones throughout the network for several malicious activities. Moreover, these clones can be re-programmed for internally attacks, such as black-hole and warmhole attack, DoS attack, extract data from the network, injecting false data disconnect the legitimate nodes through voting schemes, etc. Cloning attacks are therefore considered extremely effective because cloned devices with real information are considered to be real devices that can expose different protocols and sensor applications. Therefore, this study presented a novel technique called Hybrid Random Walk assisted Zone-Based (HRWZ) protocol for detection and revoking these clone nodes as soon as possible.

The likely results of distributed CRW based techniques RM, LSM and RAWL were chosen for the comparisons. RM has been proposed to overcome the problems of DM, however, RM entails high cost of communication where each neighbor sending $O(\sqrt{n})$ messages to reach common witness noses.

The LSM reduced the cost of communication against RM, however, LSM suffered from crowded centered problems. In crowded centered problems, the witness nodes distribution are uneven and a wide range of witness nodes are selecting from the network center for high probability of intersection on line segment forwarding path, which in results the energy of nodes are rapidly exhausting and also these nodes can become the point of interest for adversaries which in turn makes it easier to compromise. Afterwards, RAWL has been proposed for to overcome these problems of RM and LSM, However, RAWL trade in increasing the cost of communications (double of LSM) and the cost of computation for stronger security.

Furthermore, for the selection of witnesses RAWL employ simple or pure random walk which is problematic because it revisits the past nodes repeatedly which in results the nodes may die

very soon. This also reduces the likelihood of witness-node overlap, reducing the likelihood of clone detection. Therefore, this research work has dealt with the above-mentioned problems in the existing witness nodes schemes. The main purpose of this research was to develop distributed technique based on witness nodes for clone detection in static WSNs. The following steps have been taken to achieve the goals.

- Techniques based on distributed witness nodes have identified the ideal requirements and challenges that are essential to finding clones more effectively.
- Some notable limitations in existing CRW-based techniques have been identified and these drawbacks have been overcome. A distributed technique has been proposed that follows the CRW framework which detects clones with a high probability of detection and moderate overhead.

7.1 Thesis contribution

To achieve the objectives of this research work, the bellow contributions are done:

- To overcome the problems of current CRW-based schemes, this thesis presented a distributed technique called Hybrid Random Walk assisted Zone-Based Protocol.
- HRWZ combines the idea of single stage memory random walk (SSRAND) with the network division into zones. By initiating SSRAND in different zones is detecting clone nodes in the network. SSRAND is a kind of walk where a node can be visit on condition that node should not be the passed previous or current node. The needed walk steps for HRWZ are analyzed which shows that to achieve the similar detection probability of RAWL, the proposed HRWZ scheme requires less number of walk steps.
- Increase the clone detection probability while keeping communication and memory overheads
-

7.2 Limitations

In conclusion, throughout the research conduction, the following limitation has been observed which is related to the clone attack.

- Unable to detect masked replication attack (a clone node whose all neighbors have been compromised). Masked replication attack is a type of node replication attack in which

the neighbors of a witness node are compromised and they will not forward the location claim to witness nodes and thus the clone will remain undetected.

7.3 Future research directions.

- In this thesis we focused on static WSNs, and addressed the issue of cloning, future research work is required to address the issue of cloning and replication in mobile Wireless Sensor Networks and Internet of Things (IoTs). Optimization of the existing approaches is also a challenging research domain, we suggest optimization techniques using genetic algorithms can be helpful to address clone detection and replication attacks.

Appendix A

Simulator Used

.Net/Vitual Studio

Appendix B

List of Publications

Journal Papers

1. **M. Numan,** F. Subhan, W. Z. Khan, S. Hakak, S. Haider, G. T. Reddy, A. Jolfaei, and M. Alazab, *''A systematic review on clone node detection in static wireless sensor networks,''* IEEE Access, vol. 8, pp. 65450–65461, 2020.
2. Subhan, F., Ahmed, S., Haider, S., Saleem, S., Khan, A., Ahmed, S., & **Numan, M.** (2019). *Hybrid Indoor Position Estimation using K-NN and MinMax. KSII Transactions on Internet & Information Systems*, *13*(9).

Conference Paper

1. **Numan, M.**, Subhan, F., Khan, W. Z., Assiri, B., & Armi, N. (2018, November). Well-Organized Bully Leader Election Algorithm for Distributed System. In *2018 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET)* (pp. 5-10). IEEE.

References

[1] Kurniawan, A., *Introduction to Wireless Sensor Networks*, in *Practical Contiki-NG*. 2018, Springer. p. 1-46.

[2] Brooks, R., et al., *On the detection of clones in sensor networks using random key predistribution.* IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2007. **37**(6): p. 1246-1258.

[3] Akyildiz, I.F., et al., *A survey on sensor networks.* IEEE Communications magazine, 2002. **40**(8): p. 102-114.

[4] Arampatzis, T., J. Lygeros, and S. Manesis. *A survey of applications of wireless sensors and wireless sensor networks*. in *Proceedings of the 2005 IEEE International Symposium on, Mediterrean Conference on Control and Automation Intelligent Control, 2005.* 2005. IEEE.

[5] Perrig, A., J. Stankovic, and D. Wagner, *Security in wireless sensor networks.* 2004.

[6] Yu, C.-M., C.-S. Lu, and S.-Y. Kuo, *Compressed sensing-based clone identification in sensor networks.* IEEE Transactions on Wireless Communications, 2016. **15**(4): p. 3071-3084.

[7] Karlof, C. and D. Wagner. *Secure routing in wireless sensor networks: Attacks and countermeasures*. in *Proceedings of the First IEEE International Workshop on Sensor Network Protocols and Applications, 2003.* 2003. IEEE.

[8] Parno, B., et al. *Secure sensor network routing: A clean-slate approach*. in *Proceedings of the 2006 ACM CoNEXT conference*. 2006. ACM.

[9] Lee, S., *An interleaved hop-by-hop authentication scheme for filtering of injected false data in sensor networks.* 2006.

[10] Ye, F., et al., *Statistical en-route filtering of injected false data in sensor networks.* IEEE Journal on Selected Areas in Communications, 2005. **23**(4): p. 839-850.

[11] Yu, L. and J. Li. *Grouping-based resilient statistical en-route filtering for sensor networks*. in *IEEE INFOCOM 2009*. 2009. IEEE.

[12] Chan, H., A. Perrig, and D. Song. *Secure hierarchical in-network aggregation in sensor networks*. in *Proceedings of the 13th ACM conference on Computer and communications security*. 2006. ACM.

[13] Deng, J., R. Han, and S. Mishra. *Security support for in-network processing in wireless sensor networks*. in *Proceedings of the 1st ACM workshop on Security of ad hoc and sensor networks*. 2003. ACM.

[14] Przydatek, B., D. Song, and A. Perrig. *SIA: Secure information aggregation in sensor networks*. in *Proceedings of the 1st international conference on Embedded networked sensor systems*. 2003. ACM.

[15] Yang, Y., et al., *SDAP: A secure hop-by-hop data aggregation protocol for sensor networks.* ACM Transactions on Information and System Security (TISSEC), 2008. **11**(4): p. 18.

[16] Capkun, S. and J.-P. Hubaux, *Secure positioning in wireless networks.* IEEE Journal on Selected Areas in Communications, 2006. **24**(2): p. 221-232.

[17] Ganeriwal, S., et al. *Secure time synchronization service for sensor networks*. in *Proceedings of the 4th ACM workshop on Wireless security*. 2005. ACM.

[18] Hu, X., T. Park, and K.G. Shin. *Attack-tolerant time-synchronization in wireless sensor networks*. in *IEEE INFOCOM 2008-The 27th Conference on Computer Communications*. 2008. IEEE.
[19] Li, Z., et al. *Robust statistical methods for securing wireless localization in sensor networks*. in *Proceedings of the 4th international symposium on Information processing in sensor networks*. 2005. IEEE Press.
[20] Liu, D., P. Ning, and W.K. Du. *Attack-resistant location estimation in sensor networks*. in *Proceedings of the 4th international symposium on Information processing in sensor networks*. 2005. IEEE Press.
[21] Song, H., S. Zhu, and G. Cao, *Attack-resilient time synchronization for wireless sensor networks.* Ad Hoc Networks, 2007. **5**(1): p. 112-125.
[22] Sun, K., P. Ning, and C. Wang. *TinySeRSync: secure and resilient time synchronization in wireless sensor networks*. in *Proceedings of the 13th ACM conference on Computer and communications security*. 2006. ACM.
[23] Douceur, J.R. *The sybil attack*. in *International workshop on peer-to-peer systems*. 2002. Springer.
[24] Newsome, J., et al. *The sybil attack in sensor networks: analysis & defenses*. in *Third international symposium on information processing in sensor networks, 2004. IPSN 2004*. 2004. IEEE.
[25] Demirbas, M. and Y. Song. *An RSSI-based scheme for sybil attack detection in wireless sensor networks*. in *Proceedings of the 2006 International Symposium on on World of Wireless, Mobile and Multimedia Networks*. 2006. IEEE Computer Society.
[26] Chan, H., A. Perrig, and D. Song. *Random key predistribution schemes for sensor networks*. in *IEEE symposium on security and privacy*. 2003. Berkeley, California.
[27] Conti, M., R. Di Pietro, and L.V. Mancini. *Secure cooperative channel establishment in wireless sensor networks*. in *Fourth Annual IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOMW'06)*. 2006. IEEE.
[28] Conti, M., R. Di Pietro, and L.V. Mancini, *Ecce: Enhanced cooperative channel establishment for secure pair-wise communication in wireless sensor networks.* Ad Hoc Networks, 2007. **5**(1): p. 49-62.
[29] Di Pietro, R., L.V. Mancini, and A. Mei, *Energy efficient node-to-node authentication and communication confidentiality in wireless sensor networks.* Wireless Networks, 2006. **12**(6): p. 709-721.
[30] Di Pietro, R., et al. *Sensor networks that are provably resilient*. in *2006 Securecomm and Workshops*. 2006. IEEE.
[31] Yu, B. and B. Xiao. *Detecting selective forwarding attacks in wireless sensor networks*. in *Proceedings 20th IEEE international parallel & distributed processing symposium*. 2006. IEEE.
[32] Choi, H., S. Zhu, and T.F. La Porta. *SET: Detecting node clones in sensor networks*. in *2007 Third International Conference on Security and Privacy in Communications Networks and the Workshops-SecureComm 2007*. 2007. IEEE.
[33] Thakur, S.G. *CINORA: cell based identification of node replication attack in wireless sensor networks*. in *Proceedings of the IEEE International Conference on Communications Systems (ICCS'08)*. 2008.
[34] Al-Karaki, J.N. and A.E. Kamal, *Routing techniques in wireless sensor networks: a survey.* IEEE wireless communications, 2004. **11**(6): p. 6-28.

[35] Aioffi, W., G. Mateus, and F. Quintao. *Optimization issues and algorithms for wireless sensor networks with mobile sink.* in *International network optimization conference*. 2007.
[36] Akyildiz, I.F. and I.H. Kasimoglu, *Wireless sensor and actor networks: research challenges.* Ad hoc networks, 2004. **2**(4): p. 351-367.
[37] Hartung, C., J. Balasalle, and R. Han, *Node compromise in sensor networks: The need for secure systems.* Department of Computer Science University of Colorado at Boulder, 2005.
[38] Karlof, C., N. Sastry, and D. Wagner. *TinySec: a link layer security architecture for wireless sensor networks*. in *Proceedings of the 2nd international conference on Embedded networked sensor systems*. 2004. ACM.
[39] Perrig, A., et al., *SPINS: Security protocols for sensor networks.* Wireless networks, 2002. **8**(5): p. 521-534.
[40] Zhu, S., S. Setia, and S. Jajodia, *LEAP+: Efficient security mechanisms for large-scale distributed sensor networks.* ACM Transactions on Sensor Networks (TOSN), 2006. **2**(4): p. 500-528.
[41] Zhu, W.T., et al., *Detecting node replication attacks in wireless sensor networks: a survey.* Journal of Network and Computer Applications, 2012. **35**(3): p. 1022-1034.
[42] Hu, F. and N.K. Sharma, *Security considerations in ad hoc sensor networks.* Ad Hoc Networks, 2005. **3**(1): p. 69-89.
[43] Shaukat, H.R., et al., *Node replication attacks in mobile wireless sensor network: a survey.* International Journal of Distributed Sensor Networks, 2014. **10**(12): p. 402541.
[44] Akyildiz, I.F., et al., *Wireless sensor networks: a survey.* Computer networks, 2002. **38**(4): p. 393-422.
[45] Hussain, S. and M.S. Rahman. *Using received signal strength indicator to detect node replacement and replication attacks in wireless sensor networks*. in *Data Mining, Intrusion Detection, Information Security and Assurance, and Data Networks Security 2009*. 2009. International Society for Optics and Photonics.
[46] Conti, M., et al., *Distributed detection of clone attacks in wireless sensor networks.* IEEE transactions on dependable and secure computing, 2010. **8**(5): p. 685-698.
[47] Conti, M., et al. *A randomized, efficient, and distributed protocol for the detection of node replication attacks in wireless sensor networks*. in *Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing*. 2007. ACM.
[48] Yu, C.-M., C.-S. Lu, and S.-Y. Kuo. *CSI: compressed sensing-based clone identification in sensor networks*. in *2012 IEEE International Conference on Pervasive Computing and Communications Workshops*. 2012. IEEE.
[49] Kenaza, T., O.N. Hamoud, and N. Nouali-Taboudjemat, *Efficient centralized approach to prevent from replication attack in wireless sensor networks.* Security and Communication Networks, 2015. **8**(2): p. 220-231.
[50] Maheswari, P.U. and P.G. Kumar, *Dynamic detection and prevention of clone attack in wireless sensor networks.* Wireless Personal Communications, 2017. **94**(4): p. 2043-2054.
[51] Xing, K., et al. *Real-time detection of clone attacks in wireless sensor networks*. in *2008 The 28th International Conference on Distributed Computing Systems*. 2008. IEEE.

[52] Xing, K., et al. *Superimposed code based channel assignment in multi-radio multi-channel wireless mesh networks*. in *Proceedings of the 13th annual ACM international conference on Mobile computing and networking*. 2007. ACM.
[53] Naruephiphat, W., Y. Ji, and C. Charnsripinyo. *An area-based approach for node replica detection in wireless sensor networks*. in *2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications*. 2012. IEEE.
[54] Znaidi, W., M. Minier, and S. Ubéda, *Hierarchical node replication attacks detection in wireless sensor networks.* International Journal of Distributed Sensor Networks, 2013. **9**(4): p. 745069.
[55] Xia, D. and N. Vlajic. *Near-optimal node clustering in wireless sensor networks for environment monitoring*. in *21st International Conference on Advanced Information Networking and Applications (AINA'07)*. 2007. IEEE.
[56] Sujihelen, L., C. Jayakumar, and C. Senthilsingh, *SEC Approach for Detecting Node Replication Attacks in Static Wireless Sensor Networks.* Journal of Electrical Engineering & Technology, 2018. **13**(6): p. 2447-2455.
[57] Mishra, A.K. and A.K. Turuk, *A zone-based node replica detection scheme for wireless sensor networks.* Wireless personal communications, 2013. **69**(2): p. 601-621.
[58] Abinaya, P. and C. Geetha. *Dynamic detection of node replication attacks using X-RED in wireless sensor networks*. in *International Conference on Information Communication and Embedded Systems (ICICES2014)*. 2014. IEEE.
[59] Parno, B., A. Perrig, and V. Gligor. *Distributed detection of node replication attacks in sensor networks*. in *IEEE Symposium on Security and Privacy*. 2005. Oakland, CA, USA.
[60] Zhu, B., et al. *Efficient distributed detection of node replication attacks in sensor networks*. in *Twenty-Third Annual Computer Security Applications Conference (ACSAC 2007)*. 2007. IEEE.
[61] Zhu, B., et al., *Localized multicast: efficient and distributed replica detection in large-scale sensor networks.* IEEE Transactions on Mobile Computing, 2010. **9**(7): p. 913-926.
[62] Zhang, M., et al. *Memory efficient protocols for detecting node replication attacks in wireless sensor networks*. in *2009 17th IEEE International Conference on Network Protocols*. 2009. IEEE.
[63] Li, Z. and G. Gong. *Randomly directed exploration: An efficient node clone detection protocol in wireless sensor networks*. in *2009 IEEE 6th International Conference on Mobile Adhoc and Sensor Systems*. 2009. IEEE.
[64] Kim, C., et al., *A resilient and efficient replication attack detection scheme for wireless sensor networks.* IEICE transactions on information and systems, 2009. **92**(7): p. 1479-1483.
[65] Zeng, Y., et al., *Random-walk based approach to detect clone attacks in wireless sensor networks.* IEEE Journal on selected areas in communications, 2010. **28**(5): p. 677-691.
[66] Meng, X., K. Lin, and K. Li. *A note-based randomized and distributed protocol for detecting node replication attacks in wireless sensor networks*. in *International Conference on Algorithms and Architectures for Parallel Processing*. 2010. Springer.
[67] Li, Z. and G. Gong, *On the node clone detection in wireless sensor networks.* IEEE/ACM transactions on networking, 2013. **21**(6): p. 1799-1811.
[68] Zhou, Y., et al., *An energy-efficient random verification protocol for the detection of node clone attacks in wireless sensor networks.* EURASIP Journal on Wireless Communications and Networking, 2014. **2014**(1): p. 163.

[69] Khan, W.Z., M.Y. Aalsalem, and N. Saad, *Distributed clone detection in static wireless sensor networks: random walk with network division.* PloS one, 2015. **10**(5): p. e0123069.
[70] Aalsalem, M.Y., et al., *A new random walk for replica detection in WSNs.* PloS one, 2016. **11**(7): p. e0158072.
[71] Zheng, Z., et al., *Energy and memory efficient clone detection in wireless sensor networks.* IEEE Transactions on Mobile Computing, 2015. **15**(5): p. 1130-1143.
[72] Cynthia, J.S. and D.S. Punithavathani, *Clone Attack Detection Using Pair Access Witness Selection Technique.* International Journal of Computer Networks and Applications (IJCNA), 2016. **3**(5): p. 118-128.
[73] Kumar, D.R. and A. Shanmugam, *A Hyper Heuristic Localization Based Cloned Node Detection Technique Using GSA Based Simulated Annealing in Sensor Networks*, in *Cognitive Computing for Big Data Systems Over IoT*. 2018, Springer. p. 307-335.
[74] Bekara, C. and M. Laurent. *Defending against nodes replication attacks on wireless sensor networks*. 2007.
[75] Bekara, C. and M. Laurent-Maknavicius. *A new protocol for securing wireless sensor networks against nodes replication attacks*. in *Third IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMOB 2007)*. 2007. IEEE.
[76] Ho, J.-W., et al., *Distributed detection of replica node attacks with group deployment knowledge in wireless sensor networks.* Ad Hoc Networks, 2009. **7**(8): p. 1476-1488.
[77] Sei, Y. and S. Honiden. *Distributed detection of node replication attacks resilient to many compromised nodes in wireless sensor networks*. in *Proceedings of the 4th Annual International Conference on Wireless Internet*. 2008. ICST (Institute for Computer Sciences, Social-Informatics and ….
[78] Ko, L.-C., H.-Y. Chen, and G.-R. Lin. *A neighbor-based detection scheme for wireless sensor networks against node replication attacks*. in *2009 International Conference on Ultra Modern Telecommunications & Workshops*. 2009. IEEE.
[79] Cheng, G., et al. *Replication attack detection with monitor nodes in clustered wireless sensor networks*. in *2015 IEEE 34th International Performance Computing and Communications Conference (IPCCC)*. 2015. IEEE.
[80] Dong, M., et al., *LSCD: A low-storage clone detection protocol for cyber-physical systems.* IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2016. **35**(5): p. 712-723.
[81] Amudha, G. and P. Narayanasamy, *Distributed location and trust based replica detection in wireless sensor networks.* Wireless Personal Communications, 2018. **102**(4): p. 3303-3321.
[82] Pan, F., et al., *Clone detection based on physical layer reputation for proximity service.* IEEE Access, 2018. **7**: p. 3948-3957.
[83] Kitchenham, B. and S. Charters, *Guidelines for performing systematic literature reviews in software engineering.* 2007.
[84] Salam, M. and S.U. Khan, *Systematic Literature Review Protocol for Green Software Multi-sourcing with Preliminary Results.* Proc. Pak. Acad. Sci, 2015. **52**: p. 285-300.
[85] Yaseen, M., S.U. Khan, and A.U. Alam, *Software Multi-Sourcing Risks Management From Vendor's Perspective: A Systematic Literature Review Protocol.* Gomal University Journal of Research, 2013. **29**(2).

[86] Staples, M. and M. Niazi, *Experiences using systematic review guidelines.* Journal of Systems and Software, 2007. **80**(9): p. 1425-1437.
[87] Niazi, M., et al., *Establishing trust in offshore software outsourcing relationships: an exploratory study using a systematic literature review.* IET software, 2013. **7**(5): p. 283-293.
[88] Ilyas, M. and S.U. Khan. *Software integration in global software development: Success factors for GSD vendors*. in *2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*. 2015. IEEE.
[89] Alsaqaf, W., M. Daneva, and R. Wieringa. *Quality requirements in large-scale distributed agile projects–a systematic literature review*. in *International Working Conference on Requirements Engineering: Foundation for Software Quality*. 2017. Springer.
[90] Mahmood, S., et al. *Identifying the factors that influence task allocation in global software development: preliminary results*. in *Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering*. 2015. ACM.
[91] Khan, S.U., M. Niazi, and R. Ahmad, *Barriers in the selection of offshore software development outsourcing vendors: An exploratory study using a systematic literature review.* Information and Software Technology, 2011. **53**(7): p. 693-706.
[92] Bonaci, T., L. Bushnell, and R. Poovendran. *Node capture attacks in wireless sensor networks: A system theoretic approach*. in *49th IEEE Conference on Decision and Control (CDC)*. 2010. IEEE.
[93] Mishra, A.K. and A.K. Turuk. *Adversary information gathering model for node capture attack in wireless sensor networks*. in *2011 international conference on devices and communications (ICDeCom)*. 2011. IEEE.
[94] Tague, P. and R. Poovendran. *Modeling node capture attacks in wireless sensor networks*. in *2008 46th Annual Allerton Conference on Communication, Control, and Computing*. 2008. IEEE.
[95] Lin, C. and G. Wu, *Enhancing the attacking efficiency of the node capture attack in WSN: a matrix approach.* The Journal of Supercomputing, 2013. **66**(2): p. 989-1007.
[96] Angelopoulos, C.-M., et al. *Coverage-adaptive random walks for fast sensory data collection*. in *International Conference on Ad-Hoc Networks and Wireless*. 2010. Springer.
[97] Lima, L. and J. Barros. *Random walks on sensor networks*. in *2007 5th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks and Workshops*. 2007. IEEE.
[98] Alanyali, M., V. Saligrama, and O. Sava. *A random-walk model for distributed computation in energy-limited network*. in *In Proc. of 1st Workshop on Information Theory and its Application (San Diego, 2006)*. 2006.
[99] Avin, C. and C. Brito. *Efficient and robust query processing in dynamic environments using random walk techniques*. in *Proceedings of the 3rd international symposium on Information processing in sensor networks*. 2004. ACM.
[100] Gkantsidis, C., M. Mihail, and A. Saberi. *Random walks in peer-to-peer networks*. in *IEEE INFOCOM 2004*. 2004. IEEE.

[101] Servetto, S.D. and G. Barrenechea. *Constrained random walks on random graphs: routing algorithms for large scale wireless sensor networks*. in *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*. 2002. ACM.
[102] Avin, C. and B. Krishnamachari. *The power of choice in random walks: An empirical study*. in *Proceedings of the 9th ACM international symposium on Modeling analysis and simulation of wireless and mobile systems*. 2006. ACM.
[103] Lee, C.-H. *Smart sleep: sleep more to reduce delay in duty-cycled wireless sensor networks*. in *2011 Proceedings IEEE INFOCOM*. 2011. IEEE.
[104] Berenbrink, P., et al. *Speeding up random walks with neighborhood exploration*. in *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*. 2010. Society for Industrial and Applied Mathematics.
[105] Zuniga, M., C. Avin, and M. Hauswirth. *Querying dynamic wireless sensor networks with non-revisiting random walks*. in *European Conference on Wireless Sensor Networks*. 2010. Springer.
[106] Bahl, P., et al., *RADAR: An in-building RF-based user location and tracking system.* 2000.
[107] Boukerche, A., et al., *Localization systems for wireless sensor networks.* IEEE wireless Communications, 2007. **14**(6): p. 6-12.
[108] Bulusu, N., J. Heidemann, and D. Estrin, *GPS-less low-cost outdoor localization for very small devices.* IEEE personal communications, 2000. **7**(5): p. 28-34.
[109] Priyantha, N.B., A. Chakraborty, and H. Balakrishnan. *The cricket location-support system*. in *Proceedings of the 6th annual international conference on Mobile computing and networking*. 2000. ACM.
[110] Rabaey, C.S.J. and K. Langendoen. *Robust positioning algorithms for distributed ad-hoc wireless sensor networks*. in *USENIX technical annual conference*. 2002.
[111] Ramadurai, V. and M.L. Sichitiu. *Localization in Wireless Sensor Networks: A Probabilistic Approach*. in *International conference on wireless networks*. 2003.
[112] Savarese, C., J.M. Rabaey, and J. Beutel. *Location in distributed ad-hoc wireless sensor networks*. in *2001 IEEE international conference on acoustics, speech, and signal processing. proceedings (Cat. No. 01CH37221)*. 2001. IEEE.
[113] Savvides, A., C.-C. Han, and M.B. Strivastava. *Dynamic fine-grained localization in ad-hoc networks of sensors*. in *Proceedings of the 7th annual international conference on Mobile computing and networking*. 2001. ACM.
[114] Sichitiu, M.L. and V. Ramadurai. *Localization of wireless sensor networks with a mobile beacon*. in *2004 IEEE international conference on mobile Ad-hoc and sensor systems (IEEE Cat. No. 04EX975)*. 2004. IEEE.
[115] Want, R., et al., *The active badge location system.* ACM Transactions on Information Systems (TOIS), 1992. **10**(1): p. 91-102.
[116] Elson, J., L. Girod, and D. Estrin, *Fine-grained network time synchronization using reference broadcasts.* ACM SIGOPS Operating Systems Review, 2002. **36**(SI): p. 147-163.
[117] Seshadri, A., et al. *SWATT: Software-based attestation for embedded devices*. in *IEEE Symposium on Security and Privacy, 2004. Proceedings. 2004*. 2004. IEEE.
[118] Deb, B., S. Bhatnagar, and B. Nath. *ReInForM: Reliable information forwarding using multiple paths in sensor networks*. in *28th Annual IEEE International Conference on Local Computer Networks, 2003. LCN'03. Proceedings.* 2003. IEEE.

[119] Ganesan, D., et al., *Highly-resilient, energy-efficient multipath routing in wireless sensor networks.* ACM SIGMOBILE Mobile Computing and Communications Review, 2001. **5**(4): p. 11-25.
[120] Cocks, C. *An identity based encryption scheme based on quadratic residues*. in *IMA International Conference on Cryptography and Coding*. 2001. Springer.
[121] Shamir, A. *Identity-based cryptosystems and signature schemes*. in *Workshop on the theory and application of cryptographic techniques*. 1984. Springer.