SECURE AND DE-DUPLICATION BASED DATA AGGREGATION IN WIRELESS BODY AREA NETWORKS

by

Iqra Sehr

Supervised By

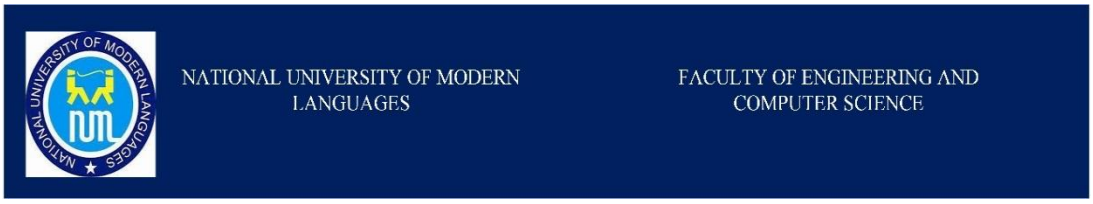
Dr. Muhammad Akbar

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

JANUARY 2018



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: SECURE AND DE-DUPLICATION BASED DATA AGGREGATION IN WIRELESS BODY AREA NETWORKS

Submitted By: Iqra Sehr Registration #: MSCS-S-16-005

Master of Science
MSCS

Computer Science
MS in Computer Science

Dr. Muhammad Akbar
Name of Research Supervisor
Signature: _____

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

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

16th January, 2019

CANDIDATE DECLARATION

I declare that this thesis entitled "*Secure and De-duplication based Data Aggregation in Wireless Body Area 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 : Iqra Sehr

Date : January 16th, 2019

ABSTRACT

Wireless Body Area Networks (WBAN) are helpful for monitoring, diagnostic, and therapeutic levels. These networks gather real time medical information by using various sensors with secure communication links. It facilitates doctors to observe a patient's health conditions by monitoring patient's vital signs away from the hospital. Sensors sense the data and forward it to the head node. The collector node consumes power to process this redundant information. It wastes too much power by sending same kind of data to next level repeatedly. During data aggregation, the collector node receives input data packets, process them and transmits it as a single packet that causes communication, energy and storage overhead.

A data de-duplication approach has been proposed to remove redundancy and ensure single instantiation of data. In this work, we have proposed a de-duplication based data aggregation mechanism that includes adaptive chunking algorithm (ACA). It identifies a cut-point between two windows. It includes fixed size and variable sized window that is identified as per minimum threshold for windows size. Our algorithm locates a second level variable length chunk based on the delimiter to improve the size of variable length window. The algorithms have been simulated using NS-2.35 on Ubuntu where TCL code is used for deploying sensing devices and message initiation. C language is used for implementing the algorithms, message receiving and sending among sensors, head nodes and sink nodes. Test results show that increase in variable sized window is measured by 65.6%, 68% and 71.2% in case of RAM, AE and proposed ACA, respectively. It results in better de-duplication identification. In this case, collector nodes consume 64% more energy as compared to sensor nodes. Results show better performance of proposed scheme over counterparts in terms of cut-point identification failure, fixed and variable length chunk size, average chunk size, number of chunks, cut-point identification failure and energy consumption.

Keywords: WBAN, De-duplication, Data Collection, Healthcare, Energy Efficiency

This thesis work is dedicated to my parents and my teachers who have helped and guided me throughout my education career that I aspire to achieve.

ACKNOWLEGEMENT

First of all, I wish to express my thanks to my Almighty Allah, who blessed me with opportunities to complete this work successfully. After that, I am thankful to my mother who encouraged and supported me in all aspects of my life. I would like to express my sincere thankfulness and gratitude to my supervisor Prof. Dr. Muhammad Akbar for his guidance and support for successful completion of this work. Yet, there were significant contributors for my attained success and I cannot forget their input, especially Asst. Prof. Dr. Ata Ullah who encouraged and guided me in a number of research activities during my thesis.

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 FIGURES

LIST OF ABBREVIATIONS

ACA	-	Adaptive Chunking Algorithm
AE	-	Asymmetric Extremum
AODV	-	Ad-hoc On-demand Distance Vector
CDC	-	Content-Defined Chunking
CDF	-	Cumulative Distribution Function
CTF	-	Collection Tree Protocol
DCC	-	Data Collection Centre
DDD	-	De-duplicated Data Dissemination
DER	-	Duplicate Elimination Ratio
EAD	-	Elasticity Aware De-duplication
FastCDC	-	Fast and efficient Content-Defined Chunking
FBC	-	Frequency Based Chunking
FSC	-	Fix-Size Chunking
HBC	-	Human Body Communications
H-IoT	-	Health Internet-of-Things
ICTs	-	Information and Communication Technologies
IoT	-	Internet-of-Things
LMC	-	Local Maximum Chunking
NLP	-	Natural Language Processing
PDR	-	Packet Delivery Ratio
PHS	-	Personalized Healthcare Systems
RAM	-	Rapid Asymmetric Maximum
SDM	-	Smart Deduplication for Mobiles
VLC	-	Variable Length Count
WBAN	-	Wireless Body Area Network

CHAPTER 1

INTRODUCTION

1.1 Overview

In this chapter, main concepts and basic definition of WBAN are discussed along with application scenarios. In this case, data is collected from medical sensors and the collector nodes aggregate data and transmit it to medical servers or central repositories at cloud. A number of challenges are also explored during aggregated data exchange and identifying duplicated data. After that, problem statement is stated followed by thesis organization.

1.2 Wireless Body Area Network (WBAN)

Wireless Body Area Network (WBAN) comprises of small sensing devices attached with human body, to monitor and exchange the data about health parameters like heartbeat, Oxygen intake level, temperature and blood pressure. Small sensors with lower resources are used to sense these health parameters [1]. They work as detectors to measure physical quantities which are converted into a signal. To reduce cost of healthcare services, WBAN are deployed where sensors are attached on wearables like clothes, bands, watch and shoes [2] [3] for continuous patient monitoring along with their ongoing normal life activities, WBAN are used to provide healthcare to elderly people, employees or patients by transmitting patient information to a storage server as illustrated in figure 1.1.

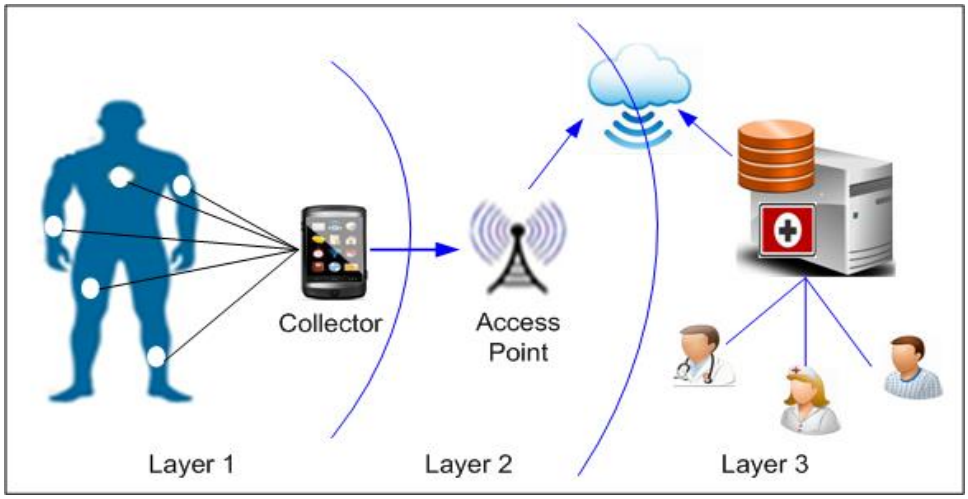


Fig. 1.1. Data Collection and Transmission from Sensors to Medical Storage Server

Heartbeat, blood pressure ECG, motion, EMG and temperature readings are generated by sensors deployed in a WBAN. All sensors are linked to a collector node placed at the center of the patient ‘s body. Real time readings generated by sensors are passed to a collector node placed at center of the body. Data aggregation and data de-duplication is applied at collector node. The collector node sends data variations and compressed data to patient’s smart phone which is further connected to the medical server of hospital. Doctors can access the data from the central medical server to monitor patient’s health. Sensors have a non-rechargeable and irreplaceable battery so their battery life issues are crucial. Abdul and Parmanand presented a good comparative study of different energy efficient routing techniques and compressive sensing approaches to enhance life time of the network [4].

Data Aggregation is the process to collect and formulate information of a specific area of interest to get its clear picture. It enables persons and organizations enable to get answer to relevant questions, analyze results and make future decisions. To maintain accuracy of research accurate data collection is required. It plays a vital role in decisions in business and enhance QoS. For example, in sales department, it can collect data from mobile applications, website visits, loyalty programs and online surveys to know customers’ demands. Data aggregation is a technique which collect and aggregate data in an energy efficient manner by using data aggregation algorithm and then send this aggregated data to base station. Sensors gathers large volumes of data, whereas they have less power so it is crucial to reduce energy consumption at

sensor level so data aggregation have a huge impact on the performance of WBAN. Centralized approach, in-network aggregation, without size reduction, cluster based approach and tree based approaches have been discussed for data aggregation. A comparative overview of various data aggregation techniques based on these strategies has been presented in [5].

De-duplication of data involves to identify the redundancy and then perform aggregation to adjust it as single instance of data. In some case, data compression techniques are further applied. Although data de-duplication is also a sub-type of compression by identifying the long similar patterns that are replaced with small sized parameters. During cyclic healthcare data dissemination, a large amount of data is shared by small sensing devices. In this case, most of the redundant data is also shared repeatedly like temperate BP remains almost same for long durations and same data for some health parameters is shared. It can also create a bottle neck, that can be resolved using some caching mechanisms. De-duplication may occur in two ways; a) Post-process De-duplication: In this method, new data is stored on a device and then it is analyzed to find duplications. Here there is no need to wait for the hash calculations and lookup to be completed before storing the data. It ensures that performance of storage operations is not degraded. But the drawback with this method is, it stores duplicate data for a short time which is an issue if the storage capacity of the system is very small. b) In-line De-duplication: In this process the de-duplication hash calculations are carried out on the target device as the data enters the device in real time. If the device finds a block that is already stored on the system, it does not store the new block and simply references to the existing block. The benefit of in-line de-duplication over post-process de-duplication is that it requires less storage space as data is not duplicated. Here the drawback is that hash calculations and lookup takes longer processing time. It means the data storage in the device can be slower thereby reducing the backup throughput of the device.

This section also presents a comprehensive overview of different data aggregation schemes that transmit patient's data to a medical server. Sensors gather large amounts of data which consume a lot of energy by sending redundant data again and again. To identify the efforts for energy efficiency during data aggregation, a number of schemes are further explored to highlight the importance of de-duplication to reduce the redundancy in aggregated data for WBAN scenario. De-duplication can be performed before transmission at collector node and after receiving at medical

server for storage as well to reduce the transmission and storage costs respectively. It also compares all the schemes for data collection in WBAN. Moreover, different research challenges for the WBAN are also explored.

1.3 Applications of WBAN

WBAN applications extend from healthcare to entertainment, sports and military. Applications of WBAN in healthcare are divided into two broad categories "Medical" and "non-medical" [5].

1.3.1 Healthcare Services for Patients

It is the most propitious field for using WBAN. Biomedical signals are collected remotely and continuously by using sensors which are wearable or implanted in the body [6] [7]. Proactive fatal and anomalies diagnosis become possible by continual monitoring of vital signs like brain and heart activity surveillance. By using previous medical profiles WBAN applications help to alert medical personnel before such cases occur to ensure public safety. Health care systems are expected to be more efficient by managing illness and reaction to crisis timely via WBAN deployment. There are following three categories of WBAN healthcare applications.

1.3.1.1 Medical Applications

Medical applications of WBAN provide continual monitoring of heart rate, body temperature and blood pressure etc. [8]. The data can be sent through a collector node which acts as a gateway to Server. WBAN is considered to be a key for early diagnosis of many serious diseases like hyper-tension, cancer and diabetes. Robots are also there to provide medical services to elderly people. Older people need more medical assistance than the young ones. They are mainly facing some major issues in routine life due to physical decline, decline in normal brain functions with other health management issues and may be the most important of all is psychosocial conditions [6]. These challenges must be taken seriously like the decline of physical condition

may lead to the inability of movement. Robots have been developed to help in health service but are still facing various challenges like acceptance of revolution in technology. Existing works have explored detailed perspective for old people health issues along with future work directions.

(i) Wearable applications: Healthcare wearable applications include measurement of heart rate, blood pressure, temperature, EEG, ECG [2] and glucose sensors to gather real time information. Wearable sensors play a vital role to improve the standard of human lives are presented in [8]. Observations of the physical phenomena can guide us to change our routine activities and tasks to prevent hazardous situations that can be avoided by spending little time in exercise or other precautionary measures.

(ii) Implant applications: Sensor nodes are implanted under the skin or in the blood stream to control diabetes, cardiovascular and cancer detection. This type of WBAN sensors are doing a good job but at the same time facing a large number of issues are still not fully resolved. Like due to constant monitoring heat generations is a problem which can lead to the thermal impairment of the human tissues. To solve the problem, the root cause is tried to avoid in [9]. Strategy was to avoid the congestion and hotspot avoidance for the traffic generated by the biosensors.

(iii) Military based medical applications are used to estimate soldiers' fatigue and battle readiness. Sensors surrounding firefighters and soldiers can predict critical situation by monitoring the level of air toxins. WBAN has critical applications in military, defense department and in many emergency response services. The individual's performance is measured due to the application of WBAN. Sensors can also increase the performance of a squad or the organization. Critical lifesaving applications are there to alert from real time threats in the battle fields by utilizing the sensors. It helps to collect the sensed data from a safe distance using wireless exchange [10]. It can also result in sensing bottleneck when a large number of sensors are exchanging information with central repositories in a continuous manner.

1.3.1.2 Non-Medical Applications

H-IoT for e-health has great potential to improve quality of life, other than medical applications there are several fields of life that used WBAN sensors to enhance the performance. In this section, some of the application domains like motion sensors, application of sensors in industry especially for the safety purpose like detection of harmful toxic gasses, smoke detection, boiler temperature detection and threshold values alerts. Sensors have also been used in secure authentication by identifying the right of access to some particular data/ information or access to a building and in many more real life applications. Wearable sensors can be used in a large variety of applications for following purposes;

(i) Motion Sensors: These applications are used to gather, detect, capture and recognize body movements and send alerts to the owner of applications. Case studies with outstanding percentage of various aspect are conducted in [11]. The values presented are taken from most of the people that are communicating online and in their response with routine usage of at least one senor at work. These sensors are capable of prediction and acknowledgment of the risk factors at work place. So application of health sensors at work place also strengthen the safety measures of workers.

(ii) Industry is the backbone of economic stability of any country. Enhanced capabilities of the IoT sensors bring it to industry like pressure sensors. Pressure sensor are useful in the industry [12]. Usually their job is to transduce the electrical signal from the readings of pressure meters. Robotics is not left without applying the IoT based sensor in industry. Pressure sensors are categorized based on the different criterion like (<10 kPa) and (10-100kPa) are low and medium pressure regimes. Depending upon kPa value they have different applications. The field of robotics heavily applies sensors to show mechanical or application of artificial intelligence which is a well-studied field of computer science. Robotic arms have their application in industry and some of them are implied in the daily routines like cooking foods, cutting vegetables and fruits, lifting items and small level parts manufacturing.

(iii) Secure Authentication: Other than medical application of sensor they are applied in other numerous fields. These applications use physical and biometric applications such as fingerprints and facial patterns. Retinal verification is very common where sensors are offering incredible services in this perspective of authentication. Privacy of personal information uploaded at cloud is one of the major concerns so by whatever means verifiable access is allowed to the individuals or organizations to protect external or private networks.

(iv) Non-medical military based WBAN application contain off body sensors that are used for emergencies. These sensors are capable of detection like detecting poisonous gas or fire in the organizations. It can timely alert the rescue team about alarming situations. Moreover, sensors can play a vital role in fire boundary detection where safe path finding is quite challenging to rescue the people or office staff. In the similar applications, the actuators are mounted at the ceilings to detect fire or smoke and timely sprinkle water to extinguish fire and generate alert messages for the rescue team as well.

1.3.2 Sports and Entertainment Application Scenarios

To maintain fitness of players, applications can maintain the record of blood pressure, temperature and heartbeat. This record can be used for safety and future training [6] [7]. For better coaching, WBAN sensors can play an excellent role. Fitness and skill level of athletes needs to be at its best for which very close observation of the movement of the players must be observed for which these sensors can offer a great job. By this, level of players and games will be at the best. This application domain may face the challenges in communication for which an approach in [13] is presented to find the best place of placing the sensors on the player's body. Other than this, it can also be used in various ways to improve the health conditions of sportsmen. WBAN sensors can sense the readings during different sleeping patterns It helps to identify the sleeping disorder faced by players which also leads to many other physical and psychological diseases. For any athlete, stamina is the basic requirement, in this regard sensors can be used to monitor breathing of players, muscles' movement and

different postures during the play and many more aspects can be covered. Entertainment applications involve process of image and mobility scenario capturing, along with post production task where actors are performing the role of some other objects. Similar applications are used for analyzing the accuracy of exercises by following right angles [7]. It also helps to save the existing exercise patters and the gradual improvements as per time. There are three types of such entertainment applications [8].

(i) Real-time Streaming: Many applications involve audio or video streaming during sensing operations like habitat monitoring. Other applications involve the voice based communication for information sharing or broadcasting. This types may have applications at airports, bus stations etc. It may also include conference calls which is mostly available in the call application software. All these types need real time streaming.

(ii) Consumer Electronics: Sensors can be deployed in microphones, MP3 players, cameras and other electrical appliances. WBAN sensors have wide application scenarios. Many sensors like heat sensor, humidity sensor and fire alarm in house are installed to have full time surveillance. On the occurrence of the specific activity sensors can automatically inform first aid management authorities. Other than this advancement in expression detection techniques have widely utilized the services of these sensors.

(iii) Gaming: Games are a big source of entertainment like virtual reality based games are played. This may include hand gesture detecting devices to move the object on the screen for some specific activity. It enables the new players to practice in a realistic environment where sensors can identify the position of a player during game practice.

1.4 Problem Statement

Sensors generate a large amount of data continuously. Head node is flooded with redundant data. From collector node, the data is transmitted to the server. The large amount of data transmission consumes battery power and reduces the network life time. Moreover, storage space is also wasted by saving same data in the server again and again. The limitation of battery life and mobility of body sensors further aggravates the problem of efficient data transmission. To reduce network congestion and to minimize packet delay it is crucial to remove redundant data by using chunking based data de-duplication technique. The main problem in the chunking schemes is that variable sized window may be smaller in size that may affect the average chunk size. Moreover, in some scenarios LMB is not found by mismatching the selection criteria. It can result in halting condition or inability to start next window.

1.5 Thesis Organization

The rest of thesis is organized in six chapters: Chapter 2 discusses about various data collection schemes in WBAN some of them incorporate deduplication in it. Chapter 3 explains the methodology of our research work. Chapter 4 describes the proposed dynamic data de-duplication along with aggregation for data exchange in WBAN. An Adaptive Chunking Algorithm is proposed to identify cut-point during chunking mechanism in de-duplication. In Chapter 5, results and analysis have been discussed where simulation environment is also explored. Chapter 6 discusses about conclusion, achievements and some possible future work.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

In this chapter, literature review of related data de-duplication based schemes are explored. WBAN architecture and challenges are also explored along with application scenarios for data collection and transmission to central repositories. Next, data de-duplication based schemes are investigated that include the data custodian based approaches for central and distributed data storage. After that, data chunking based schemes are also studied that present the content based chunking where cut-point identification is the key mechanism. It provides the basis for evaluating the pros and cons of existing schemes and identifying the research problem in cut-point identification mechanism.

2.2 WBAN Architectures

WBAN consists of small sensors for heart pulse, BP and ECG monitoring [6] [7]. These sensors are attached over the body or implanted inside the body of a patient, to perform scheduled monitoring and report whenever a threshold value is crossed. Battery power of these tiny devices is very limited that hinders the smooth operations for longer duration and reduces the lifetime of the node and network as well. These are detectors to measure physical quantity and converted these into electrical signal. It also involves the cost factor that should be reduced to enhance the applicability of body area sensors at large scale. Therefore, WBAN are deployed in an easy attach and detach manner on clothes, shoes, hair pins and friendship bands etc. [8]. The applications of WBAN are demanding and reliable, used greatly in Medical, Electronics, Gaming and

entertainment, fitness sports and agriculture [4]. Figure 2.1 shows that data from body sensors is collected and forwarded to personal mobile. After that, the aggregated data is transmitted to access point that further transmits it to medical history database server via internet. Server delivers instantaneous information to family of patients, doctors or other medical staff.

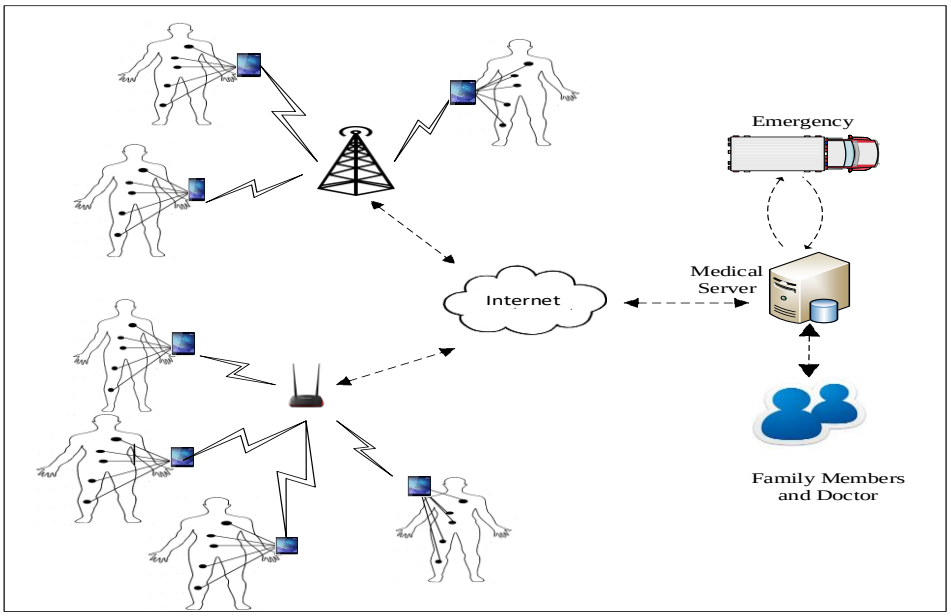


Fig. 2.1. Data Collection and Transmission in WBAN

Information and Communication Technologies (ICTs) have promising potential for health internet-of-thing. Detailed study from various aspect of ICTs including taxonomies and challenges yet to be addressed can be found in [16]. It provides a deep insight considering ICT paradigm. It is not limited to the cloud based services only, rather expended to the FOG, mobile edge, cloud computing, mobile cloud, IoT and M2M. Healthcare system have been found very useful and widely studied by the researchers. It also consists of WBAN which is the life line of today's ubiquitous health paradigm. Different from the traditional healthcare systems WBAN have some advantages like: Effective and consistent monitoring of the patients or elderly people. Capturing the physical phenomena remained the outstanding benefit of the WBAN. Quick response is another promising feature which can save precious human lives by timely informing the nearby health service centers like hospitals. In some cases, patient don't need to visit the hospitals frequently because physicians at the hospital can

continuously observe the patient's physical situation. Necessary measures can be taken by timely availability of lifesaving drugs and access to ambulance services. Power consumption is not a big issue because these health sensors do not need to perform computations and also do not need to transmit sensing reports at large distance due to availability of local storage centers. A device can be placed nearby or attached to a patient's body or implanted within patient's body. These devices can exchange data with collector node to aggregate the observed physical condition and then forward it to a local center. This transmission can occur with the help GSM, or Wi-Fi access points. In case of Wi-Fi services, it is more device friendly as less power is to be consumed.

WBAN is a cloud-based service where a medical server is available in the architecture. Since servers are powerful devices there is no issue of computational power and the storage capacity. This is an edge of Cloud based services of WBAN. To show typical application scenario of the proposed work, medical server is deployed but its application can be extended to other real life applications. In Section 2.3, variety of its application from various perspectives are presented. WBAN is already being applied and found to enhance services in much better than ever been in the past.

Observation of the tiny health sensors is not just useful typically for one person. Patients' generated data can also be mined by the medical researchers to provide better and efficient services to the patients especially in emergency situations where quick response is required. Medical practitioners can improve their vision of the patient's case study, so WBAN is a kind of social welfare service as well. Moreover, it also leads to the personalized e-health services like the genetics of every person is not same. Due to this, the human body responds differently, varying from person to person. Proliferation of H-IoT devices has done a remarkable job in this regards too as the people are different, may have unique environmental conditions (extremely hot or cold areas/weather). WBAN also have great potential to help the paradigm of medication to develop drugs based on factors mentioned earlier. To reach up to goal of personalized healthcare systems (PHS), it must overcome various challenges. Some of the challenges are pointed out in [17] like the PHS sensors should be cost effective and should provide accurate physiological readings of a patient's condition. This can help doctors to make a good medical prescription as per the need of a particular patient. Another supportive situation that can be helpful for an educated decision can be through the available family

history of diseases. WBAN has the potential to achieve PHSs when some basic challenges are met.

WBAN system can be sub divided into four basic stages; i) Layer of sensing devices which actually observe the phenomena. This can be any sensing devices from the paradigm of IoT; ii) Sensing report is transmitted to the internet which is to be delivered to the network layer; iii) Transmitted data is processed for data mining to extract useful knowledge for mapping to current condition of the patient. It can be a turning point for effective medial response; iv) Finally, the application layer is there to offer services [17].

Once patients' data is transmitted to the data centers at cloud it must be protected for privacy even when to be used for study purposes, permission must be taken from the patients. Securing the personal data especially, for the case of PHS is one of the main requirements that must be fulfilled for the large scale application [18]. This also demands for light weight IoT based security mechanisms for secure transmission to avoid active and passive attacks. ZigBee 802.11.5 and Wi-Fi 802.11 are the most commonly used wireless transmission techniques to transmit data from sensors, which are more vulnerable to attacks due to the wireless nature. Selection of the transmission technique may depend on various aspects like the communication range, data transmission rate and required battery life time. Wi-Fi based technique is more suitable when high data transmission rate is required. But ZigBee approach consumes less battery power and has range up to 10 Km. Automated responses to the sensor reports are desirable but it demands human like intelligence. H-IoT based sensors are divided in categories like activity sensors which monitor activity like the gyroscope which measures healthy walk as prescribed by the medical experts of 10,000 steps per day. These kind of sensors can help to change our unhealthy life style to schedule and well calculated one so that it can regulate the daily routine. Sensors that can observe physiological conditions are also there to provide real time services. Sensors can measure blood pressure, heartbeat, breathing rate, electrocardiography (ECG), electroencephalography (EEG), and other quantities depending on the type of sensor [19]. Advancement in H-IoT gadgets have reduced need for expensive healthcare facilities for constant monitoring of the patients. Home environment is more suitable for psychological patients which can be achieved through the remote sensing facilities.

Power of the IoT based health sensors still faces same challenges as these sensors need to communicate which may reduce their power then observation will not be possible. Data is taken frequently from patients which is to be delivered to the backend server from where physicians can access the updated situation of patients with access privileges. Transmission of readings to the data centers requires energy so there is a need for efficient techniques which can save transmission cost. This will also increase the battery time of sensing devices. This is also important because batteries are not very convenient to be replaced and are not economical as well.

2.3 Challenges of WBAN

WBAN is special type of wireless sensor network (WSN) so challenges of WSN also exist for WBAN. Latre et al. presents a survey on WBAN schemes. The main objective WBAN is to achieve a fault tolerant arrangement of sensors on human body with minimum delay with highest level of throughput and with lowest power consumption. Social and ethical requirements of user's like safety, privacy and ease of use and reliability entails so much importance. There are many challenges which are to be addressed by the research community in future. Some of these challenges are discussed in this section.

2.3.1 Interoperability

The data is to be exchanged across different technologies like Zig Bee and Wi-Fi but these technologies are different in configuration, data exchange protocols and compression mechanisms. Moreover, different frequency, data rate and bandwidth are supported. The system should be capable of managing the connection during communication switching from one network to other and bridging mechanism should be smoothly applied.

2.3.2 Energy Consumption

WBAN consists of small nodes with small batteries which add power constraints in communication. It is considered as one of the major problems and challenges and needs investigations and solutions. In WBAN, frequent replacement of nodes must be

avoided especially when nodes are implanted within the human body because they are not easily accessible. Because of this, one has to consider the tradeoff between the energy capacity and energy consumed by the processing and communication operations in order to use energy efficiently. Reliability is also challenging due to low transmission power and small sized antenna of wireless sensor devices. It affects signal to noise ratio that causes a higher bit error rate and decreases the reliable coverage area. However, reliable transfer of data in WBANs and medical monitoring systems is crucial. Reliability in data delivery in a network can be measured by Packet Delivery Ratio (PDR) and Bit Error Rate (BER). PDR represents the ratio of the number of packets received by the receiver to the number of packets generated by the sender, while BER represents the ratio of the number of error bits to the number of bits generated by the sender. The reliable data transmission should be investigated for low power body sensor networks which is still a challenge for BASNs.

2.3.3 Node Heterogeneity

Sensor nodes are most likely heterogeneous [9] so to prevent the formation of hotspots, the nodes must be employed such that energy of each node in the network is consumed homogenously. Specific applications of WBANs may require heterogeneous data collection from different sensors with different sampling rates. Therefore, QoS support in WBANs may be quite challenging. Maintenance of data integrity while working with wide range of devices during data collection of WBAN on different layers is difficult.

2.3.4 Interference

There are chances of collision and packet loss due to mutually interfering wireless signals. There must be a way so that nodes will recognize that they belong to which network so, that minimum interference will occur. WBAN performance can degrade significantly due to precarious signal integrity. Various schemes have been proposed to mitigate this issues in WBAN like. Interference-Aware Channel Switching, Lightweight and Robust Interference Mitigation Scheme and many others [10].

2.3.5 Data Aggregation

Data aggregation involves the collection of data from different sensing devices but it is transmitted by using a single device. It receives the data from neighboring nodes and transmits single large aggregated message. Aggregation helps in reducing communication cost and energy consumption because message transmission consumes more energy as compared to computation.

2.3.6 Data Security

WBAN are naturally dynamic, attacker may impersonate the credible sensors and get the patient data easily. Moreover, medical information can be eavesdropped, injected and modified. So, data transmissions should be handled securely at sensor and collector level.

2.3.7 Data Integrity

It should also be ensured that the data is not altered during transmission. An attacker can modify the data packets and reconstruct the message on same format to falsify the data. As the sensed data is private and any change in it may be dangerous for the patient. So data must be accurate and reach in time to the destination to ensure timely treatment for patient. To solve the problem of path loss and message integrity hash function [11] can be deployed in WBAN architecture.

2.3.8 Reliability

High level of reliability is required for WBAN as it directly affects the healthcare of patients. Medical professionals should receive correct data to meet user requirements.

2.4 De-Duplicated Data Aggregation in WBAN

Literature have been studied for the purpose of collecting the patients' data from various healthcare services. Background knowledge remained the focus of study by

considering the health perspective, aggregating data and its transmission to the Data Collection Center (DCC) for various applications. In this section, secure collection of data and the perspective of de-duplication are explored.

Exchange of sensing data is done by proposing a cluster-based model by Youssef et al. in [4]. Inaccurate positive values are reduced for the collected sensed data, in turn this improved the accuracy of proposed scheme. Sipal et al. studied fitness scenarios at three hub locations in [12]. The design is centered to waist-centric, network centric to a head and other to a footwear. For every hub location four spots on the body are considered; back, chest and both arms. H(f) is a function representing the transfer through wireless channels among the hub locations. Readings are taken and analyzed for two activities; push-ups and for the squat exercises. These activates are conducted for male and female genders. Authors also conducted a study by considering the cumulative distribution function.

Muhammad Quwaider proposed a scheme by considering the processing on the collected data and scalable storage in [13]. Broadband infrastructure like 3G and LTE is required for the collection of healthcare related data. Sensed data is transmitted to the servers at cloud on hour bases with the rate of 360 packets by using Wi-Fi which further perform processing to fetch useable knowledge. At first, Bluetooth is used for the collection of data which is then sent to the sink node and then to the smart phones or personal digital assistants (PDAs) by Wi-Fi and at last cloud is the final destination for further evaluation on sensed data. Ad-hoc On-demand Distance Vector (AODV) protocol is used by priority queue formula to differentiate the data traffic. Result of simulation also include the effects of priority and priority less queue. AODV lacked due to the exploitation of FIFO planning approach especially when emergency case is considered. Pre and non-pre-emption conditions are included to overcome the drawback in AODV [8]. For healthcare monitoring, radio frequency (RF) and non-RF techniques of communications are utilized [14]. Human body communication is used for non-RF. It also implied capacitive coupling or galvanic coupling. Capacitive coupling is based on only one electrode in receiver and transmitter while alternatives remained floating. In contrast to this galvanic coupling have electrodes at both ends. RF and HBC are compared based on applications. The concept of smart WBAN is based on heterogeneous method with multi-radio. WBAN hub is responsible for the communication among devices like relay or bridge and may have distinct radio standards. Model consists of three major components first one is BAN, secondly, Node

and third one is relevant to the data processing and management. A special band is dedicated to the smart BAN so that accessed by the authorized users. Data and control channels are used separately. Due to its features smart BAN is utilized for special applications or for the observation of special phenomena.

Quilling Tang proposed AODV which is a routing protocol [15] on the base of priority (urgent or routine) data to decrease the loss of packets at the same time to reduce transmission latency. Three tier architecture is considered consisting of various sensors of ECG, EMG, motion, hearing and blood pressure sensors. These body sensors are connected to a head node. This node transmits the collected data to smart phone or to a nursing station which monitor the sensed data, this is the third level. Ultimately the collected data is sent to the database server for the purpose of keeping the record. If any emergency situation occurs alarm is triggered. Every sensor node have a threshold value. Collected data is transmitted to phone which the send replica to main server. Real time transmission of data is offered by the scheme when exceeds a threshold, transmission is made to sink node in short time to avoid loss in terms of time. Scheme bring a good packet delivery percentage with high throughput.

Youfeng chiu et al. introduced Elasticity Aware De-duplication (EAD) [16], this scheme allows user to specify a trigger $T(0,1)$ value for migration which represents accepted deduplication level of that particular user. Some amount of RAM is assigned for indexing before the process of de-duplication. This works on sampling by the case analyses for efficient adjustments. Analysis of result from experiments verify EAD improved system efficiency up to four times. Scheme shows 98% duplicate data by consuming 5% memory.

The very initial purpose of introducing the backup storage is to decrease the cost of storage, overhead of transmission and for the efficient use of space when wireless multimedia networks are considered. Comma separated frame $C = \{c_1, c_2, \dots, c_n\}$ is generated by Chunking Algorithm from source data of sensors. C' is $|C \cap C'| \rightarrow n(n \leq m)$ is the recently sample data, if this is similar/equal to C then it means that most elements of recently sample data are stable. Assume that collector node contains R data and multimedia sensor knows that c' which is similar to chunk contained in R. Then the contents that are duplicated should not be exchanged but the indices for data can be shared. Further stable sequences end up with improved results regarding the reduction of storage space and also for the need of network bandwidth [17].

A coordinator is included in privacy preserving de-duplication protocol to broadcast startup message to all D_i in dataset. It includes criteria for query represented as *d* along with the and the probability represented as *P*. According to figure 2.2, a data custodian leader is represented as D_L which is D_1 in this case. It can be rotated as per utilization. Next, data custodians are shown in a chain including $D_L \rightarrow D_2 \rightarrow D_3 \rightarrow \cdots D_{i+1} \rightarrow D_N$. Based on the existing protocol, a new sub protocol is developed on the basis of nearby neighboring nodes. It is utilized to eliminate the duplication along with reduction in storage overhead [18]. The system can suffer from traffic analysis based attack where an intruder can extract d_A redundant strings from any of the data custodian which was declared as corrupted due to bit alteration or communication anomaly. It can be predicted with a probability *P* as specified in equation (2.1).

$$P = (d - d_A)(1 - P(false\ positive)/|N - D_A| \tag{2.1}$$

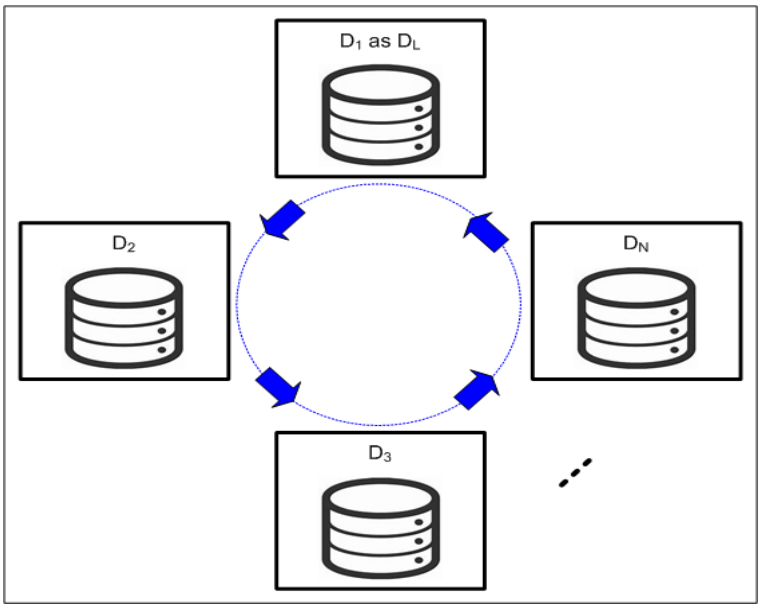


Fig. 2.2. Ring Topology in Data Custodians

2.5 Data Chunking based Schemes

Deduplication algorithm are mainly categorized in two types Fix-Size Chunking (FSC) and Content-Defined Chunking (CDC). Main function of both types of algorithms is same that is to eliminate the repeated chunks to save resources. Main issue of FSC is that revision of whole file is required even if small number of chunks are updated and boundary of all the chunks is modified [19].

2.5.1 Sliding Window based Chunking Schemes

CDC algorithm is more flexible in redefining the boundary chunks as it is based on contents and not on the size [20]. Break point satisfaction is the basis of CDC algorithm which is an important component in deduplication. This component has an important impact on the deduplication ratio and duplication performance. Sliding window based schemes have been in use for 15 years but it is less efficient as byte by byte deduplication is required. This remained the motivation to present novel and efficient solutions [21]. Leap-based CDC is proposed which is claimed to improve deduplication while deduplication ratio is also not compromised. Deduplication is generally comprised of four steps: chunking, chunk fingerprinting (compute fingerprint), fingerprint indexing, querying and data storage along with its management [21]. Major drawback of the sliding window based CDC is that it imposes predefined conditions which can cause forced breakpoints for the larger chunks of data. This can also enhance deduplication ratio. To handle this, an approach with secondary condition is introduced in [22] which works in similar way like sliding window CDC approach. The idea was that if the chunks are larger than the defined size, then secondary condition can be exploited and even if it did not satisfy the maximum limit of chunks where forced breakpoint can be exploited. The leap-based algorithm [21] also has similar concept of secondary condition and it is aimed to reduce the ratio of forced breakpoints. Rolling hash is not applicable so the pseudo-random transformation is presented as a replacement. It has presented the comparison of sliding window and leap-based CDC from the perspective of CPU and memory consumption and deduplication ratio. Results show better performance of the proposed approach.

2.5.2 NLP based schemes for Chunking

Natural Language Processing (NLP) algorithms can be used to reduce redundancy in data. It is a well-known area of study computer science especially in the paradigm of artificial intelligence. It is aimed for the study of how to program computers to process large amount of natural language data. Several methods exist as a solution to NLP problems. Winnow is a family of stat-of-the-art algorithms characterized for

robustness and filtering of irrelevant features. Winnow can handle only linearly separable data. Having the motivation of extending the application to non-linearly separable data an approach in [23] is presented. It is considered as a classification problem which can help in statistical prediction model to make prediction. Due to inherent capability of the robustness, high dimensions of data can be handled but on the other hand when considering the real life applications. It may need a lot of storage capacity which may turn this advantage into deficiency. Because of this, the winnow algorithm only considered necessary features. A highly featured "SNoW" architecture for NLP remained point of interest where irrelevant attributes of data are to be considered. CoNNL-2000 has been used for comparison as a dataset in proposed scheme. Scheme [23] considered text chunking problem as a sequential prediction problem presented in equation (2.2).

$$P(\{t_i\}|\{\omega_i\}) = \prod_{l=0}^{m} P\left(t_l \middle| \{t_j\}_{j<l}, \{\omega_i\}\right) = \prod_{l=0}^{m} P\left(t_l \middle| x_l\{t_j\}_{j<l}, \{\omega\}\right) \qquad (2.2)$$

Where,

ω_i is a sequence of text that is tokenized $\{\omega_i\} = \{i = 0,1,2,3, \ldots m\}$ is considered to be known

t_i is to be predicted on the given $\{\omega_i\}$

x_l is a "featured vector" while $\{t_j\}_{j<l}$.

Conditional probability model is exploited in this scheme. For further details about the selection of x_l section 3 of this scheme can be considered. First and second order types of chunks are considered, the difference is that first order considered necessary feature whereas second order may have high dimension which can make prediction problem more complex. So first order is adopted in by [23]. To train the vector weights proposed generalized winnow algorithm is exploited. Authors claimed 10 % reduction in errors when compared with original winnow algorithm. Errors may occur due to four main reasons; i) ambiguity in the nature of data; ii) some feature can be ambiguous so errors can occur; iii) insufficiency in the training of data; iv) utilization of not a very efficient algorithm can also be the cause of errors.

2.5.3 Fast and Efficient CDC Scheme

By criticizing CDC based approach in [21], a modified scheme titled as Fast and efficient Content-Defined Chunking (FastCDC) is presented in [24]. It is based on Gear hash based CDC where hash judgment is enhanced but simplified at the same time. For the purpose to achieve speed sub-minimum chunk points are skipped. Normalized distribution is exploited to reduce the ratio of deduplication for a specific region. The size of chunk in normalization are claimed to be large enough to handle the smallest chunk size. It categorized literature in algorithmic and hardware oriented CDC based approaches and declared hashing and hashing judgment as two stages of CDC schemes. Hashing mean assigning hash value to chunk whereas hashing judgement stands for the comparison of identified chunk-points. Selection on exploitation of Gear hash is made according to the presented comparison of Robin and Adler approaches. Gear hash requires fewer calculations as compared to Rabin.

2.5.4 Frequency based Chunking

In Frequency based Chunking for Data de-duplication, there are several platforms at cloud that are data centric and a large amount of data is generated and received which require large storage space. Saving data in storage is difficult and may be un-necessary to record as it could be redundant. For this deduplication (dedup), CDC is considered as a good solution but it reduces the size of chunk to increase dedup ratio. A scheme with the idea of removing the redundant data is proposed in [19] and Frequency Based Chunking (FBC) algorithm is presented. The basic difference from conventional CDC approach is that the proposed approach considers the frequency of data with the intention to eliminate the frequent data or more specifically metadata. Proposed FBC has applied CDC algorithm to fetch coarse grained chunks and then chunks from CDC algorithm are evaluated for the repeated data segments. An algorithm is used to identify the highly frequent segments of data named as statistical chunk frequency algorithm. Pre and parallel filtering approach is used to help in identification of global repeated chunk. It claims the increase of performance in comparison with stat-of-the-art algorithms. Pre-filtering process requires only XOR operation and for every chunk to

be in the parallel filtering stage this stage must be passed. For evaluation of FBC, Duplicate Elimination Ratio (DER) is given by equation (2.3).

$$DER = \frac{input\ stream\ length}{total\ sum\ of\ the\ sizes\ of\ chunks\ produced} \quad (2.3)$$

As one of the motivation of FBC was to eliminate the metadata, so to achieve it DER is modified to fulfill the desired objective as given in equation (2.4).

$$DER_{meta} = \frac{|S|}{|S| - Gain_A(S)} = \frac{|S|}{\sum_{c_i\ \varepsilon\ distinct(s)} |c_i| + (|index| + |Chunk\ list|)} \quad (2.4)$$

Where,

$gainA\ (s)$ is the storage space saved when algorithm A is applied on the data

S represents stream

$|S|$ is length of the stream.

2.5.5 Smart Chunking based schemes

Data deduplication is not the only but one of the mainly used technique to reduce the network traffic and storage at cloud service provider that includes mobile phones as well. It has promising application for mobile based users where speed is required. Data compression by the process of deduplication utilizes hash values to identify the similarity of data and then its removal is done. Smart Deduplication for Mobiles (SDM) is presented in [25] for low power mobile devices to save the transmission cost, which is claimed for higher accuracy level and better performance than its predecessors. It claims to be the first on mobile devices from the perspective of deduplication. SDM is designed to take the edge of multi-core architecture available in mobile devices. Motivation of the SDM is taken from various aspects like: performance limitation of mobile devices, bandwidth limitations, network tariff and limited network coverage which demand for less amount of data to be transmitted. Chunking methods in deduplication are categorized in file and block level. File have only one hash value whereas block level chunking has multiple blocks with hashing values accordingly. One edge of block level chunking is that it is more resistant to change than in file level as one small change in a chunk will not affect the whole file. In SDM, selection between

these methods is decided based on the time that would be required to de-duplicate the data. One unique feature of the SDM approach is that it is autonomous in the selection of configuration of file type. More importantly it has negligible overhead in comparison with the time saved for uploading.

2.5.6 Sub-Chunk De-duplication based Schemes

CDC algorithm is criticized for computational complexity as analyzed by [25], where CDC based hashing algorithm consumes more processing time during de-duplication. Later on an approach in [20] is presented which did not used hashing mechanism for chunking known as Rapid Asymmetric Maximum (RAM), bytes value is used to cut points instead of hashing. To handle this, RAM exploited fixed and variable sized window to search for the maximum value byte which served as the cut point of the chunk. It leads to make less comparisons which can save computational power. More than this RAM is able to attain the CDC property of algorithm.

An Anchor-Driven Subchunk Deduplication approach in [26] is presented by taking the motivation form [27] and [28]. The former is "Bimodal" approach which deal with the smaller chunk size and the later one "Fingerdiff algorithm" with large size. In [27], large size is totally ignored whereas due to only consideration of large size it may not be suitable for main memory which will definitely affect the performance of the system. In [26], an anchor based sub-chunking scheme is the simplified form of HYDRAstor [29]. It has two steps in the process of deduplication one is for large and other is for sub-chunks. Mapping of sub-chunks to container context is exploited in this anchor-driven approach. It requires small memory so suitable for main memory processing. Anchor based scheme is suitable both for the small as well as for large sizes. Its duplicate elimination ratio shows improvement as compared to its predecessors.

Romanski et al. introduced dynamically perfected sub chunk de-duplication for a limited size [30]. In [31], the concept of HBC was introduced. The proposed scheme achieves improved bit error and data-rate at the same time security as well in comparison with RF communications. Gnawali et al [26] proposed frequently used Collection Tree Protocol (CTP) as a reference protocol to perform data collection in

WSNs. The basic theme of CTP is to construct one or more collection trees in such a way that every one of them is rooted at a sink. It transmits and receives data by exploiting these trees. Sensing nodes in a tree utilize stop-and-wait ARQ with at most the maximum of M (configurable) retransmissions. Each node sends its data along with the data of other nodes. Ghamari et al. [27] proposed a scheme with the objective to minimize required energy of transmission in WBANs by exploiting energy harvesting techniques along with low-power MAC protocols. Other than these measures, application layers also adopted a better sampling and data transmitting strategy which is in accordance with the application. Sipal et al. [28] proposed a technique for hub location in human body at head, foot and waist. The wireless channels measure the frequency between the hubs and four nodes (chest, back, and upper arms). Proposed framework results show that overall best performance is for a hub located on the temple and the worst overall performance is achieved for a hub on the waist.

Bjorner et al. proposed a fixed size sliding window where local maximum bytes are identified. It can be identified after a fixed size in between two text windows extracted from input string. Beside this, it consumes more processing period and also increases the computation costs. It may not be appropriate for de-duplication process executed for the data related to health parameters because human lives depends on efficiency of scheme [32]. In [23], another chunking algorithm involves the cut-point that is located at the right side of a fixed side window similar to LMC. The scheme is titled as Asymmetric Extremum (AE) because it finds a maximum or extreme value located at central point between two windows. The key difference is that, its variable sized window is at left side and fixed sized window is at right side in contrast to counterparts. A mechanism for the selection of cut-point during the process of chunking is proposed in [33] known as Rapid Asymmetric Maximum (RAM). In this case, a cut-point is identified byte wise where the maximum value is identified at the edge of the variable sized chunk. It shows a combined cut-point after two consecutive windows as shown in algorithm 2.1. Hashing mechanism in excluded when cut-point is being identified. Figure 2.3 presents the cut-point identification mechanism of various approaches like; LMC, AE and RAM. In this work, the proposed model adopts fixed sized window and variable sized windows similar to RAM but delimiters are used for identifying cut-points. For the identification of cut-point, it traverses on the right side of fixed size window.

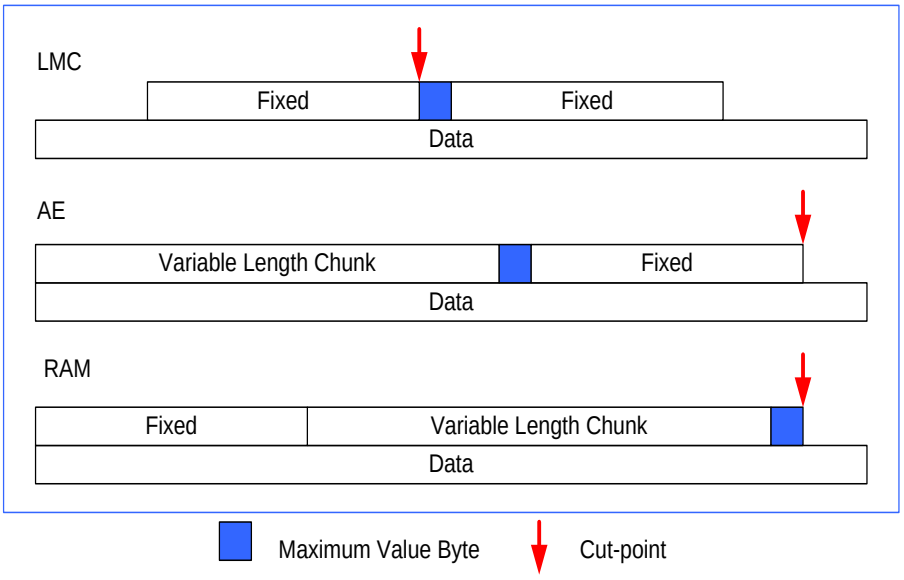


Fig. 2.3. Cut-point Identification for Chunking Mechanism

Algorithm 2.1: Algorithm for RAM based Cut-point Identification

Algorithm 1: Rapid Asymmetric Maximum Algorithm (RAM)

1. ***Input:*** *Input String Str, Length of string L*
2. ***Output:*** *cut-point i*
3. ***Predefined Values:*** *window size w*
4. ***Function*** *RAM_Chunking_CutPoint (Str, L)*
5. *Set i to 1*
6. ***While*** *i < L do*
7. ***If*** *Str[i] >= max_val* ***then***
8. ***If*** *i > w then*
9. *Return i*
10. ***End if***
11. *max_val = Str[i]*
12. *max_position = i*
13. ***End if***
14. *Incr i by 1*
15. ***end while***
16. ***end function***

2.6 Summary

Different categories of data de-duplication schemes along with chunking mechanisms have been explored in literature to cover the related schemes for our work. The main focus is on identifying the cut-point to distinguish between data chunks based on the content of acquired data. Moreover, different WBAN architectures are studied to highlight the most frequently used components and their interaction scenarios. Additionally, a number of open research challenges are also identified as per different application scenarios for WBAN.

CHAPTER 3

METHODOLOGY AND FOG ORIENTED WBAN ARCHITECTURE

3.1 Overview

In this chapter, a methodology is presented along with proposed FoG oriented WBAN architecture to discuss about key components and their functionalities. It explores the healthcare based data dissemination in FoG-oriented network setup with local storage at FoG server and central repositories at cloud. In the similar vein, a number of opportunities are explored that highlight the importance, benefits and need for the proposed architecture. Moreover, open research challenges are also identified that should be resolved to achieve the intended functionality for de-duplicated data dissemination and transmission.

3.2 Fog Oriented WBAN Architecture

This is the era of mobile devices with rich sensing capabilities. Portable health sensors are now part of our daily lives like wrist watch consisting of many sensors. In time detection of life threatening viruses like (Chikungunya virus) can save precious human lives. This virus may lead to multiple organ failure. Remote areas have less medical facilities so these are more prone to diseases. A framework in [1] is based on fog and cloud assisted architecture. It has three layers, one collects data (health/body sensors), second is fog layer which is at the edge of cloud and lastly cloud layer which

is the backbone of the architecture. This architecture is similar to our proposed model. Cloud layer is used for the storage of data that may be used by the FoG server at FoG layer in our proposed model. FoG servers are able to analyze the reported health condition of a patient and may respond to the aggregator node (AN) to guide for necessary immediate actions. The aggregator node may collect data by itself as it consists of onboard sensors or simple aggregate the sensed data by the IoT devices as shown in figure 3.1. AN also performs first level deduplication of the collected data and transmits it to the FoG server. At this stage, AN does not perform chunking on the sensed data. FoG server transmits the data to the cloud server after performing de-duplication based on the CDC scheme with chunking. As CDC is a compression technique so it removes unnecessary repeated data to save the storage space at cloud server. This action can reduce the needed bandwidth and latency as well. Cloud server may have many data repositories or data centers where data is kept and provided on demand.

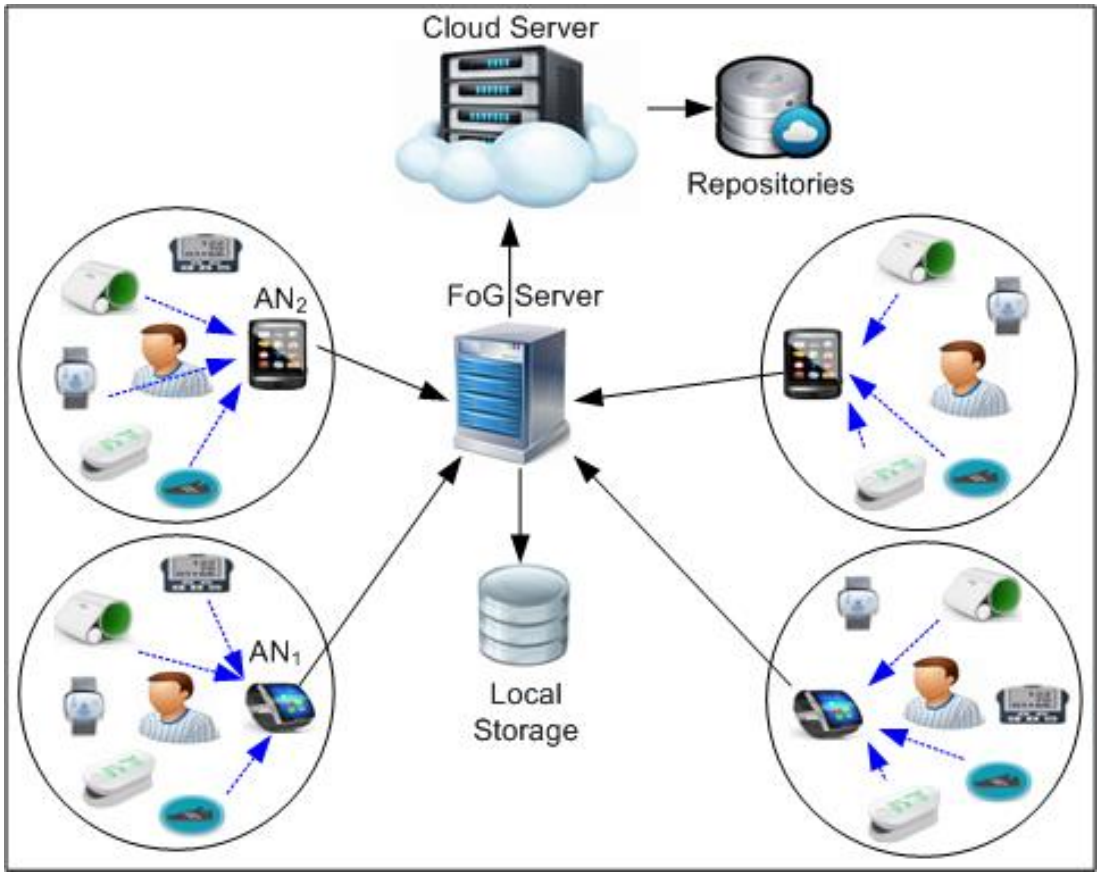


Fig. 3.1 FoG oriented De-duplicated Healthcare Data Dissemination

3.3 Proposed Architecture

Proposed model offers many applications in health sector like cooperation among physicians at different locations is improved so they can share their views on the condition of a patient. More than this proliferation of sensing IoT devices may transmit consistent reports even without the knowledge of the patient. So practitioners can take immediate action to avoid any emergency situation even before it is known the patient. Proposed FoG oriented cloud approach bring information at the fingertips of a doctor as local storage is available with FoG server. This local storage at FoG server is analogous to the cache memory which keeps most recently accessed data ready for use. It can avoid the sensing bottleneck at physical layer by caching the healthcare data for some time to avoid message congestion on the network as well.

3.3.1 Connectivity Improvement

Data is shared through local storage and if some requests cannot be fulfilled then FoG server based cloud services can be used. Cloud has a vast storage capacity that can provide almost every required information to give response to the operators of health sensor or to the sensor directly. For the case if requested information is not present at one server then request/requests can be met by getting services from other cloud service provider (CSPs). In this way availability of the information to analyze and to make a decision about a lifesaving situation is available when a serious medical condition occurs.

3.3.2 Quality of Service

Quality of services in terms of low latency rate and increased throughput is provided by the proposed model. This is possible due to the distribution of cloud capabilities near the edge nodes, where almost direct access is available. This low latency becomes more promising, because it deals with human health. In emergency cases, a real-time response is required either by physically reaching the location of patient or in the form of prescriptions and guidelines from medical experts serving at far off places.

3.3.3 Low latency

Presence of the FoG architecture is assurances of low latency as for every query cloud server is not needed to be contacted. Health services are delay sensitive and require real time response. Fog computing is comparatively more predictable service that can reduce the response time which is very important for serious medical situations. Health sensors can also play vital role to a personalized health care as medical paradigm is moving towards personalized medicines for which large analysis of symptoms may be done at FoG server. Cloud services may be used by applying deduplication techniques to reduce the response time.

3.3.4 Local Resource Management

FoG server in the proposed model, fetches computational and storage capabilities at the edge for resource constrained sensing devices. Smart devices can access resource rich repositories for deciding about health critical situations by analyzing real-time records that have been saved after de-duplication to reduce storage space and computation as well.

3.3.5 Bandwidth Management

Deduplication is an efficient technique as it effectively utilizes the assigned bandwidth. It removes repetitive data, ultimately less data is transmitted. In the proposed model deduplication is performed at two stages which makes the scheme more efficient. This can also be efficient from the perspective of data transmission cost as AN node may have to use mobile data when Wi-Fi is not available in which transmission of large and varied data sizes of chunks is not an issue. By reducing network traffic bandwidth is saved which prolongs the continuous monitoring of health sensors especially in the case of old age patients monitoring application.

3.3.6 Improved Energy Efficiency

Energy utilization is also a challenging task to make the FoG oriented architecture as green architecture. Sensors consume a continuous amount of energy,

therefore, energy efficient solutions are needed to promote green computing. Efficient energy utilization is beneficial for protecting our environment as well. Less amount of transmission means efficient utilization of energy which is one of the most important concern of the sensing devices due to availability of limited battery power. End sensing nodes do not need to transmit sensed data to FoG server directly. For this purpose, ANs collect/aggregate the sensed data from sensors and then transmit it to FoG server which ultimately is delivered to the data centers at cloud. Deduplication is also another aspect which improves intelligent and optimized use of energy.

3.3.7 Improved Services Accessibility

Utilization of FoG servers is assurance that services would be available at one step from the health sensors attached to the human body. FoG based architecture is proposed as it shifts the computation and storage capabilities from cloud to end devices. This reduces latency and improves service quality. More than this cloud computing is also available for backup services. Number of edge nodes can be increased to provide better services in urban areas which are densely populated.

3.4 Challenges for Proposed FoG Oriented Architecture

FoG based cloud model is proposed in this research work which has a lot of benefits. Proposed model has many applications but in this work it focuses on health services. Availability of health sensors improve life style of human beings by continuously monitoring the human body. It also has many important challenges that should be handled to provide efficient and dependable solutions. It is also the key concern of service users and stakeholders to ensure quality parameters for smooth operations. Challenges required for efficient and successful implementation of proposed model are discussed in this section.

3.4.1 Storage Capacity Evaluation

Storage capacity of a scheme is analyzed by evaluating the computational capabilities and data processing activities in a FoG based model. First level and second

level deduplications are applied to reduce the redundant data, which is verified by analyzing the storage capacity utilized at fog and cloud servers. In health based services timely responses are the desirable. Aggregation and transmission of sensing data form IoT sensors to local storage and repository at cloud is expected to reduce the delay due to application of CDC deduplication. Aggregation of sensed data is done to reduce transmission cost. If data collected by sensors is sent individually it would take more communication time. It also increases throughput and reduces latency. Collected data from various health sensors can be analyzed to get complete picture which is helpful to map the real condition of a person. It also opens new horizons for the researchers to develop efficient solutions that can reduce transmission delay and likely to provide real time and more authentic responses from medical experts at distance. Proliferation of the sensors may produce a lot of network traffic which must be refined before storing data in repositories to conserve storage space.

3.4.2 Patient's Mobility based Data Management

Health sensors are mostly the body sensors which are attached to the body of a people who have moved from one place to another. This mobility may cause issues like battery and coverage area. People can move in vehicle or using other means so continuous tracking of the sensing data may be a challenging task. In this situation, de-duplication can be useful to reduce the transmission rate of data as transmission is an expensive process. This situation can be worse in hilly areas where service quality may be low. In this case, more transmission power is required which may lead to a difficult situation for battery powered mobile devices. The challenging situation needs to be handled by researchers, from the perspective of more efficient transmission mechanism which can reduce the unnecessary wastage of battery power.

3.4.3 Delay Reduction

FoG server is employed in the proposed model to reduce the response time as most frequently required data and its analysis is stored at local data repository. Carry and forward approach can produce a lot of unnecessary network traffic which increases communication delay. So it is challenging that some efficient approach is used on the

middle nodes to avoid carry and forward approach. In our proposed approach filtration is done by performing the analysis on the data to generate required data only. It also provides an insight to produce more efficient schemes which can differentiate the needed and not required data in the scenario of FoG based cloud services in health scenarios. This may require buffering at the nodes exploited for transmission. Moreover, if less efficient mobile service is available then more delay will occur in the delivery of packets to FoG server. To handle these challenges, scientists must focus on efficient exploitation of FoG server based cloud services.

3.4.4 Secure Communication

Health sensors can be used to observe the state of a patient. This information may be confidential and should not be publicized. Local storage at FoG server is one of the intermediate devices which process the personalized information so privacy remains a concern of user for health related matters. Other than this, storage at cloud can also be accessed by the internal or external entities or even by a CSP. Secure and authentic access to the stored data at data centers must be ensured to ensure privacy of health related data. This can be done by employing access policies. Literature is also available on the idea of avoiding the unauthorized access from internal authorized nodes. Access must be granted based on the role or on the base of privileges and monitoring of access must also be tracked. A large number of sensors may transmit data simultaneously which must be handled independently and anonymously. Sensed data must be transmitted to the right destination which demands transmission security and data integrity. Malicious intermediate nodes may collect transmitted data from NA node to FoG server. Active attacks can alter the data in transit or passive attacks can silently listen and capture the data for launching future active attack. These attacks should be considered by the researcher while designing efficient and dependable solutions for providing secure transmission and anonymous handling of personal data. Handling of heterogeneity of data reported from various health sensors must be integrated in a universal manner which is also challenging and also requires attention. Continuous monitoring in case of severe medical conditions with rapid connectivity is also desirable. Trust of the user must be ensured. Therefore, some trust management schemes need to be developed to ensure privacy of sensitive information. Data from sensors can be used by doctors to treat in much better way as case studies would be

available. Trusted third party can be employed to ensure the anonymous use of reported cases for study purposes.

3.4.5 Dynamic Data Aggregation

Data can be shared by multiple persons in a home or a group of patients in a ward. The proposed architecture also supports that multiple patients can send their aggregated data in the form of concatenated string to the collector node that can aggregate the entire data. It can also perform the de-duplication at that time to reduce the size of message for saving communication cost and redundancy as well. Due to the involvement of cell phone, a patient's data can be shared from any place and FoG server can also receive and de-duplicate the data as well by identifying the unique identity of the patient. It also acts as a primary key to save records in the data base. It can also help to maintain a linkage between the patients of same house or same region to identify the effects of similar environment to cause similar symptoms to avoid any sort of critical diseases as a carrier.

3.5 Summary

Methodology of the proposed FoG oriented WBAN architecture has been discussed to highlight the intended functionalities of different components. The system includes local and global storage options for FoG server and cloud servers respectively. In this chapter, we have also discussed about possible opportunities that can be achieved by using the proposed architecture. Moreover, a number of open research challenges are identified to attract new researchers by adopting our proposed FAV architecture. It can achieve better functionalities for providing de-duplicated data dissemination and transmission.

CHAPTER 4

DE-DUPLICATED DATA DISSEMINATION SCHEME

4.1 Overview

In this chapter, a de-duplication based chunking scheme for healthcare data dissemination scenario is presented. Data is collected from smart healthcare devices and then aggregated to transmit to central repositories. Sink nodes perform data de-duplication to eliminate redundancy and reduce size of data actually transmitted on the network. A novel adaptive chunking algorithm is also proposed to identify a cut-point to divide a set of data into chunks which is based on the contents of data.

4.2 De-duplicated Data Dissemination Scheme

In this section, a novel De-duplicated Data Dissemination (DDD) Scheme for WBAN is presented. It utilizes the FoG server located the at the edge of the network. It helps to reduce communication delays for healthcare data exchange from sensing devices to cloud having large amount of global storage. To manage the chunks of a large sized data string, we have proposed an adaptive chunking algorithm which is applicable at sink node or FoG server.

4.2.1 Secure Data Dissemination

In this section, it has discussed about the data dissemination from the smart healthcare devices. In this scenario, a number of patients or even employees have been

considered with healthcare sensors attached on body or clothes. Sensors regularly take data and share with central repositories. It has been observed that healthcare data involves the duplicate readings like temperature values for a person does not vary rapidly. Similar is the case for other healthcare parameters like blood pressure, ECG, heart rate etc. It should be identified before sending to servers to save communication cost and also identify when there is significant change in value as per defined threshold values for each healthcare parameter. At device side, the best point is the first level collector that takes data from neighboring devices and then transmits the aggregated data towards sink or FoG servers. It should perform first level de-duplication and replace the healthcare parameter values with Boolean representation to indicate that same value holds as transmitted in last message. It will be replaced with 1-bit Boolean representative value unless new significant change is not identified in that reading. The collector node also encrypts the data using the secret key shared with sink node. At next level, de-duplication is performed at sink where a large amount of data strings are received and can be further processed using our proposed algorithm to identify data chunks for performing de-duplication. It can help in identifying large sized chunks because a large amount of data is present for a particular time period to share with the cloud servers and store at central repositories for future decision making about health conditions. Data from sink to cloud is also encrypted to guard against security attacks. Collector node also takes hash of data to guard against bit alteration attacks and ensure the message integrity. In the similar vein, timestamp values are also included in messages to guard against the replay attacks. It reduces the communication overhead and energy utilization. It also reduces communication delays from sensing devices to cloud repositories. In our work, he detailed description for security protocol is not included and considered as beyond the scope of this work which is more focused on de-duplication.

4.2.2 Adaptive Chunking Algorithm (ACA)

On receiver side, the data is received as concatenated strings where a different delimiter is used to differentiate the data of various collectors from multiple regions. Healthcare parametric value are extracted and then stored in the central repositories. From the extracted data, the identities of the source devices can be identified to store

the values as per patient or user for further analysis. It also helps to maintain a history of records and timely generate the alerts in case of alarming situations. These records can also be used by the doctors and care takers to analyze the health conditions. A list of notations is presented in table 4.1.

Table 1. List of Notations

Notation	Description
w	Window Size
i	Cut-point
Str	Input String
L	Length of String
m, n	Loop index values
VLC	Variable Length Chunk
AN	Aggregator Node / Collector
D_1	Delimiters for concatenating data from sensing nodes
D_2	Delimiters between concatenated data of multiple patients
$Str\_Sp_{L1}[]$	Split String using D1
$Str\_Sp_{L2}[]$	Split String using D2

In this scenario, it is considered that the data driven approach at sink node where content defined chunking mechanism is used for achieving de-duplication mechanism. The process begins by identifying a cut-point or pivot to first identify the fixed size window. Next, variable sized window is obtained by marking another cut-point after as threshold window size to obtain large sized chunk that can help to achieve better de-duplication. It is also dependent on the setting the acceptable threshold window size *w* and the variable length count (VLC). The proposed scheme considers the VLC to restrict the acceptable chunk size where VLC is larger than the *w*. On the contrary, if VLC_1 is smaller than *w*, then it will check for second level delimiter to grab extra bits to produce VLC_2 that can enlarge the overall length of variable sized window as shown in figure 4.1.

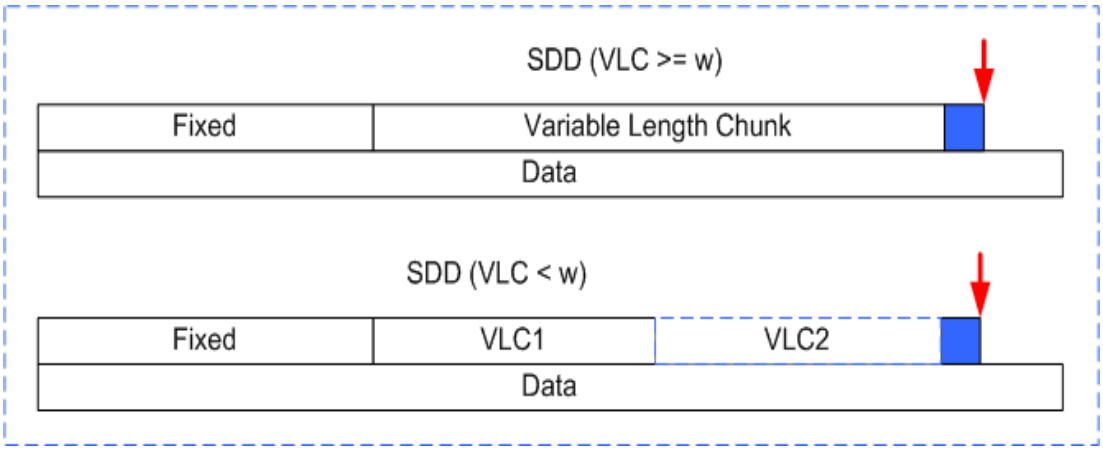


Fig. 4.1. Proposed Adaptive Chunking Mechanism using VLC

Our proposed Adaptive Chunking Algorithm (ACA) is presented in algorithm 4.1 to explore stepwise description of cut-point identification process. In this mechanism, a delimiter D_1 has been used to divide the data in chunks especially in variable sized window identification. Delimiter is used by each sensing device to aggregate data where each healthcare parameter is differentiated from other. Moreover, collector nodes also use a different delimiter to aggregate data from multiple sensing devices. Initially, a data string *Str* (stored at sink) is split into tokens and saved in a collection Str_Sp_{L1}. It first verifies that the if the first token size is less than or equal to the *max_chunk_size* then it further verifies that the chunk size $Str\_Sp_{L1}[m]$ is greater than or equal to window size *m.* If the condition is true with larger size string then it returns the size of chunk value as *size* $(Str\_Sp_{L1}[m])$. In cases where chunk size is less than *w*, then it attempts to generate a larger chunk by finding the index value of next occurrence of delimiter D_1 as *Indexof*$(Str\_Sp_{L1}[m++], D_1)$ and assigning to variable *i*. It will keep on increasing the chunk size unless greater than *w*. In cases where size of $Str\_Sp_{L1}[m]$ is greater than the *max_chunk_size*, then the level 1 string $Str\_Sp_{L1}[m]$ is further split to get level 2 string titled as $Str\_Sp_{L2}[n]$. In this case m is index value of collection $Str\_Sp_{L1}[m]$ to identify chunk at that index. Similarly, *n* is index value for $Str\_Sp_{L2}[n]$. In our proposed ACA, size of variable length widow is managed on the basis of delimiter which is placed in aggregated data to differentiate among healthcare parameter values. It reduces the cost for finding next cut-point in the data string. The algorithm's complexity is O(n) due to inclusion of one iterative operation for identifiying the index value.

Algorithm 4.1: Adaptive Chunking Algorithm (ACA)

Input: Concatenated Data String Str, Data Length L

Output: cut-point i

Predefined values: window size w, max_chunk_size

Function *DCA_Chunking_CutPoint (Str, L)*

1. *Set i to 0 and m to 0*
2. $Str\_Sp_{L1}[]$ *= Split (Str, D1)*
3. ***If*** *size (*$Str\_Sp_{L1}[m]$*) <= max_chunk_size* ***then***
4. ***If*** *size (*$Str\_Sp_{L1}[m]$*) >= w* ***then***
5. *Return i = size (*$Str\_Sp_{L1}[m]$*)*
6. ***Else***
7. ***While*** *size (*$Str\_Sp_{L1}[m]$*) < w*
8. *i = Indexof(*$Str\_Sp_{L1}[m]$*, D1)*
9. ***End While***
10. *Return i*
11. ***End if***
12. ***Else***
13. *Set n to 0*
14. ***If*** $Str\_Sp_{L1}[m]$ *> max_chunk_size* ***do***
15. $Str\_Sp_{L2}[n]$ *= Split (*$Str\_Sp_{L1}[m]$*, D2)*
16. ***If*** *size (*$Str\_Sp_{L2}[n]$*) >= w* ***then***
17. *Return i = size(*$Str\_Sp_{L2}[n]$*)*
18. ***Else***
19. *Incr n by 1*
20. *Return i = Indexof(*$Str\_Sp_{L2}[n]$*, D2)*
21. ***End if***
22. ***End if***
23. ***End if***
24. ***End function***

Figure 4.2 illustrates the flow chart for proposed adaptive chunking mechanism where cut-point is identified based on the delimiters inside the healthcare data. It help to ensure the availability of cut-point within the input string. It also ensures that the variable sized window is restricted as per maximum chunk size. Moreover, VLC should also be larger than the minimum threshold size.

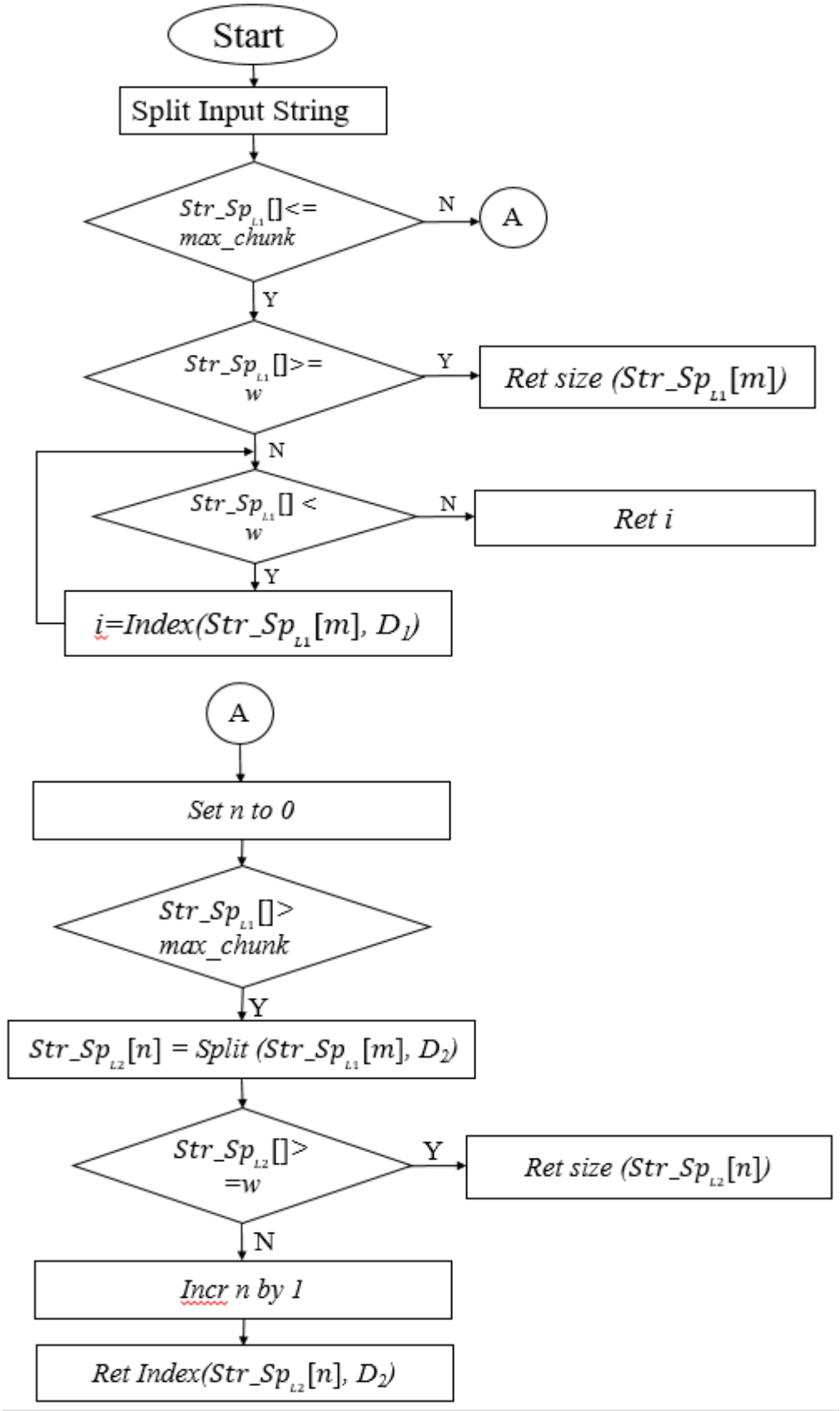


Fig. 4.2. Flow Chart for the Adoptive Chunking

4.3 Summary

Proposed scheme for de-duplicated data dissemination is presented in this chapter. In this scheme, sink nodes collect data from multiple healthcare devices and then perform aggregation to exchange a bulk of data in a single message to reduce communication cost. Our proposed work further helps to reduce the communication cost by eliminating redundant information. It also reduces the storage cost by storing less data. In this regard, a novel adaptive chunking algorithm is proposed that helps to efficiently identify the cut-points as compared to preliminaries. It divides data into chunks by using cut-points which are based on delimiters in the content of data. A flow chart is also presented to illustrate the adaptive chunking mechanism.

CHAPTER 5

RESULTS AND ANALYSIS

5.1 Overview

In this chapter, we have presented results and analysis along with simulation environment where WBAN is deployed for sensing, data collection and transmission scenarios. For the simulation, an event-based network simulator NS-2 is used where proposed ACA algorithm is implemented using C language. Moreover, RAM, AE and LMC algorithms are also implemented for comparison. A list of simulation parameters is provided to help in understanding simulation environment and configuration parameters. Results are presented for average chunk size and cut-point identification failure.

5.2 Simulation Environment

The proposed work is validated by using Visual Studio 2015 where C Sharp is used as programming language and SQL Server 2012 is used for loading DataSet for healthcare parameter values. In this case, a comma separated file containing values from dataset has been utilized to populate records in tables of database. On server side, an application has been developed in C Sharp to implement ACA, LMC, AE and RAM algorithms. Next, it evaluate the computation cost for number of chunks and average chunk sizes on the same dataset for healthcare parameters. In conjunction with this server side application, a client side android based mobile application is also developed. It is used to add more parameters in the dataset as per readings taken from patients or other data sets. It has been verified that values are entered as per minimum and maximum boundaries for healthcare parameters like temperature cannot exceed 106^{o}F. To further validate the sensors and collectors, we have also simulated using NS-2.35 on Ubuntu for the low level calculations at sensing devices during message

exchange and extract the energy consumption and residual energies. To accomplish this, separate classes have been developed to manage node configurations and functionalities for sensing devices, collector nodes and Sink nodes. TCL files are used to deploy the medical healthcare parameters sensing devices as per x, y coordinates and configure the nodes as per devices configurations. It also initiates a message exchange at a particular time that is further handled by C code files where send and receive functions are executed along with packet configuration parameters. Moreover, mobility of a patient have been considered by using setdest() function in TCL files. Finally, an energy model is also implemented to extract the residual energy after each operation during aggregation, transmission and de-duplication. It also prints the residual energy for all nodes like sensing devices and AN in the trace file that can be interpreted using AWK scripts. Results are extracted for average chunk size, fixed size and variable sized windows identification, Number of chunks probability of failure for cut-point identification and energy consumption at node and collector. List of simulation parameters is given in table 5.1.

Table 5.1. List of Simulation Parameters

Simulation period	350 Seconds
Broadcast packet size	256 Bytes
Transmission Power at Vehicle	0.819 μJ
Receiving Power	0.049 μJ
Channel Type	Wireless
Propagation Model	Two Ray
Mac Protocol Type	Mac/802-11
Queue Type	Queue/DropTail/PriQue
Antenna Type	Omni Antenna
Max Packet in Queue	50
Agent Trace	ON
Router Trace	ON
Mac Trace	OFF
Multiplier (ψ)	2 - 4
Size of Input String (Bytes)	10000 – 50000 Bytes

5.3 Average Chunk Size

During chunking in de-duplication, it is quite critical to identify right cut-point that results in improvement in terms of average chunk size. If the cut-point is near to the fixed size window, then chunk size remains smaller. In case of LMC [32], fixed size window is similar in size with the second window where cut-point is located at center of two window. It reattempts when a cut-point is not identified in first attempt that slows increases computation overhead as well. In case of AE [23], the second window is fixed in size and the first window size is varied to adjust and link with the previous chunk. It achieves larger chunk sizes. In RAM [33], cut-point is identified at the edge to get larger chunk sizes. In the proposed ACA, a minimum threshold has been applied on the second window which is variable in size. The position of cut-point is readjusted if the window size is smaller than a threshold size. Figure 5.1 shows the for average chunk size in bytes. Results show that average chunk sizes 185 bytes, 664 bytes and 670 bytes for LMC, RAM and AE, respectively. Our proposed ACA performs better by achieving an average chunk size of 678 bytes.

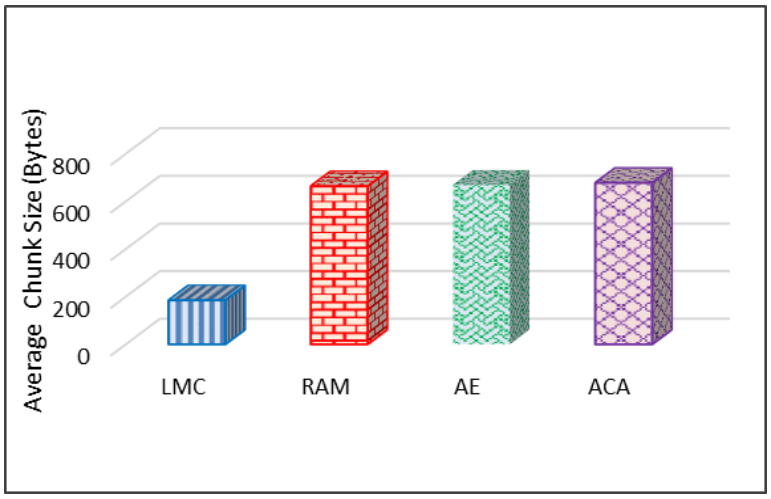


Fig. 5.1. Average Chunk Size

Figure 5.2 shows the sizes for two main windows including fixed and variable sized windows. Results show that for a fixed size window of size 250 bytes, the sizes of variable sized window are 428 bytes, 420 bytes and 414 bytes, for proposed ACA, AE and RAM, respectively. On the contrary, LMC maintains the same for two windows which is 250 bytes. Big chunk sizes are generated sue to large sized windows. It helps in achieving better redundancy elimination. Results show that variable sized window is grown by 65.6%, 68% and 71.2% in case of RAM, AE and ACA, respectively.

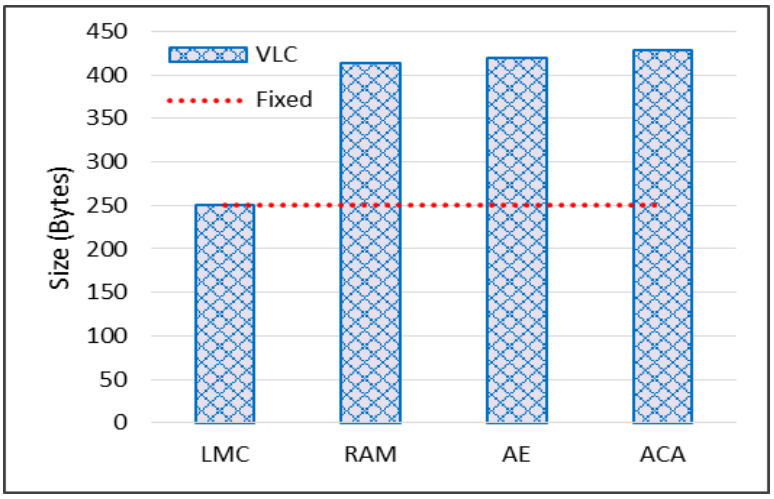


Fig. 5.2. Chunk Size Difference in Fixed and VLC Sizes

5.4 Number of Chunks

Figure 5.3 shows the number of chunks created during de-duplication for different sizes of input string varying from 10,000 Bytes to 50,000 Bytes. We have considered the average chunk size to extract the average number of chunks. Results shows that for an input string size of 30000 Bytes, LMC produces 162.1, AE creates 44.77 and RAM creates 45.18 chunks whereas the proposed ACA generates 44.24 chunks. It is observed that ACA generates less number of chunks with larger chunk sizes. It improves the chances of de-duplication identification from a large chunk.

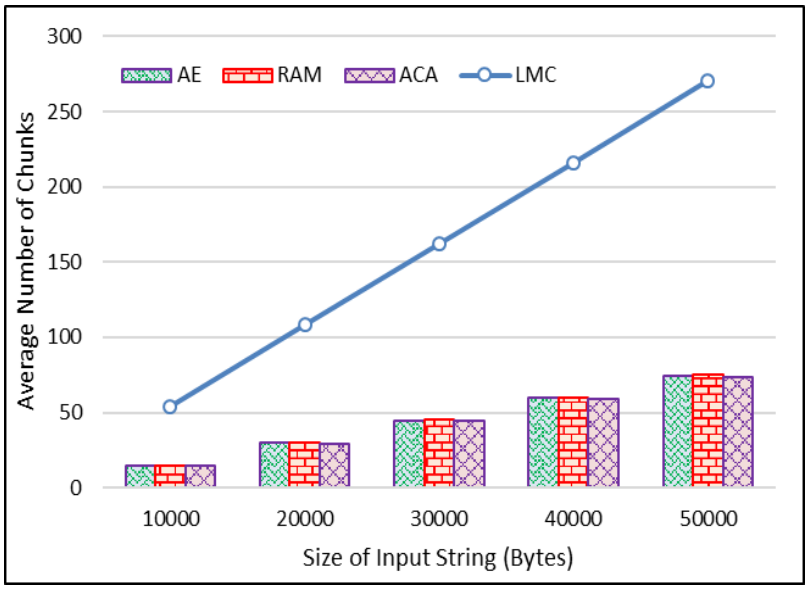


Fig. 5.3. Average Number of Chunks

5.5 Cut-point Identification Failure

Figure 5.4 shows probability of failure for identifying the best suitable cut-point during the process of chunking. It may happen that cut-point is not identified in the given input string because of mismatching the cut-point finding criteria. In this scenario, a multiplier ψ is applied as a unit to calculate the computation cost. It also helps to measure the probability. It has been observed that, LMC has most chances of failure to find cut-point. Results illustrate that for a multiplier of 2.5, LMC faces 26.6% failure whereas AE bears 4.16% failure to find cut-point. RAM and our ACA reduces chances of failure to 0.83% only.

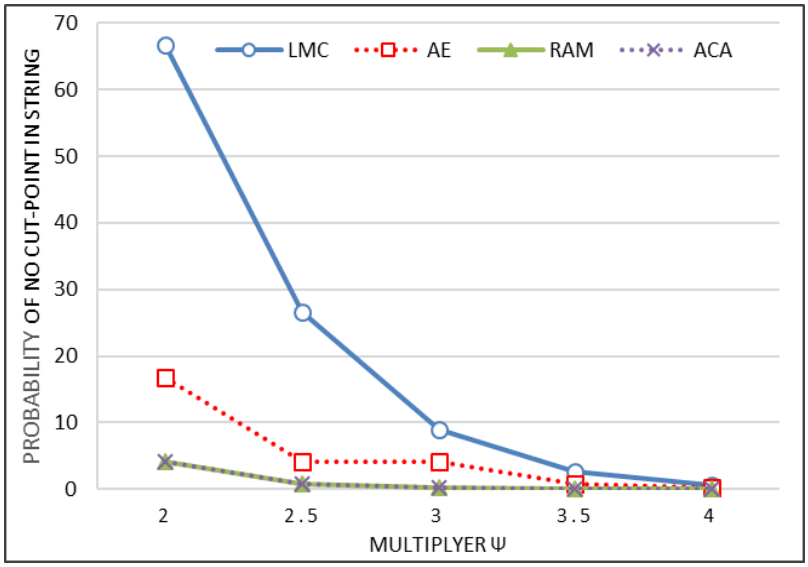


Fig. 5.4. Probability for Cut-point Identification Failure

5.6 Energy Consumption

Figure 5.5 shows the energy consumption at AN during de-duplication and chunking mechanism. Energy model is implemented in NS-2 to print the residual energy values in trace files. Next, the energy consumption is extracted by taking difference of energies as per time *t* in seconds from trace files generated after simulation. Results reveal that sensor nodes S_1, S_8 and S_{12} consume 0.003029 µJoules and S_6 consumes 0.003044 µJoules at time t=0.5 seconds. In first step, the energy

consumption values in micro joules are extracted for sensing devices and then energy consumption for ANs are extracted.

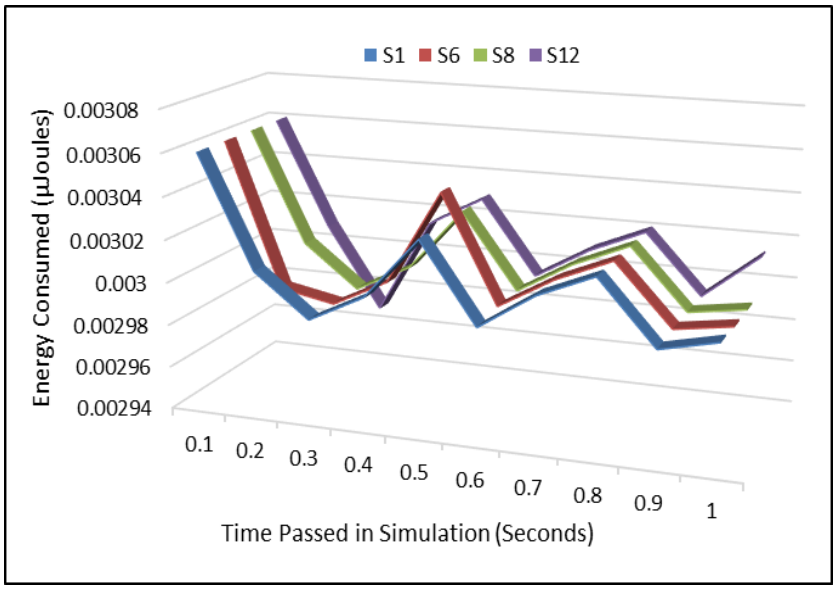


Fig. 5.5. Energy Consumption by Sensing Devices

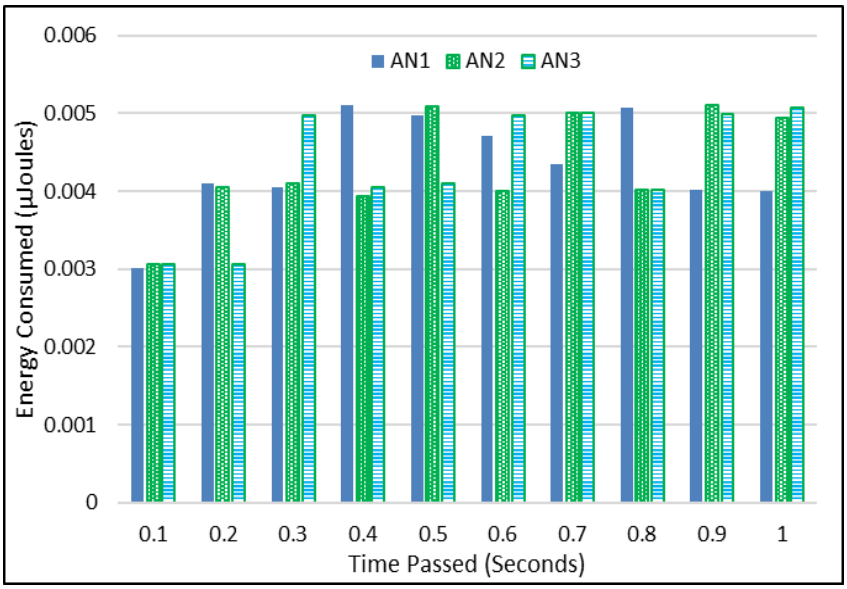


Fig. 5.6. Energy Consumption by Aggregating Devices

Figure 5.6 shows the energy consumption at AN during data aggregation and transmission for data de-duplication. Results reveal that AN_1 consumes 0.004977 µJoules, AN_2 consumes 0.005089 µJoules and AN_3 µJoules at a particular time

t=0.5. Results show that the collector nodes AN_1 and AN_2 consumes 64% more energy in contrast to a sensor node S_1, S_8 and S_{12}. Moreover, during data deduplication and data exchange AN_3 consumes 68% more energy.

5.7 Summary

Results are presented in this chapter where simulation environment is explored along with configuration parameters for wireless scenario. Moreover, parameter variations along x-axis are also included in table. In the similar vein, network model is also explored and deployment scenario is considered for WBAN sensing for healthcare. It performs data collection, aggregation, de-duplication and transmission scenarios. For the implementation of existing and proposed algorithms, we have developed codes in NS-2.35 using TCL, C and AWK scripts. Results are presented for average chunk size, number of chunks, cut-point identification failure and energy consumption at sensing devices and AN.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Overview

In this chapter, concluding remarks are presented to highlight the key contributions of this work. It also briefly describes the problem identification, proposed solution and results. Finally, the future work is discussed to elaborate the possible future extensions of this work.

6.2 Conclusion

During healthcare data de-duplication, the sink node extracts the healthcare parameters from the input concatenated strings shared by the collector nodes. These strings contain the delimiters to differentiate between the data taken from different patients or employees. These values are aggregated along with level one de-duplication to eliminate the redundant values for healthcare parameters shared by different sensing devices. In our proposed DDD protocol, it further involves the second level de-duplication at sink or FoG server using our novel algorithm ACA. It identifies a cut-point located at the center of fixed and variable sized windows. To validate our work, initially a testbed is developed to implement de-duplication algorithms at sink node for proposed and existing schemes. It involves a dataset for healthcare records which is further populated using android application. We have simulated our work using NS-2.35 for deploying sensing devices, collectors and sink nodes for healthcare data

dissemination using TCL scripts and C language code. Results show that variable sized window is grown by 65.6%, 68% and 71.2% in case of RAM, AE and proposed ACA, respectively. Results show better performance of proposed scheme over counterparts in terms of number of chunks, fixed and variable sizes for chunks, average chunk size, probability of failure to identify cut-point and energy consumption.

6.3 Achievements

In this work, the proposed scheme has achieved better results due to proposed ACA algorithm and data dissemination model. Following metrics have been improved in terms of reducing de-duplication during healthcare data exchange.

i. Communication Overhead
ii. Variation in Chunk sizes
iii. No Cut-point Identification
iv. Average Number of chunks
v. Energy consumption.

6.4 Future Work

In future, impact of de-duplication and cut-point identification at cloud servers can be explored to analyze the effects at reducing storage cost for healthcare data. Moreover, it can be evaluated that reduction in communication cost also affects the storage cost especially in healthcare data.

REFERENCES

[1] Z. Hui, G. Lijuan and L. Hui, "Secure and Privacy-Preserving Body Sensor Data Collection and Query Scheme," *sensors*, pp. 16-179, Monday 2 2016.

[2] C. Riccardo, M. Flavia and R. Ramona, "A survey on wireless body area networks: Technologies and design challenges," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1635-1657, 2014.

[3] M. B. Deena and M. Ammad-uddin, "A Survey of Challenges and Applications of Wireless Body Area Network (WBAN) and Role of A Virtual Doctor Server in Existing Architecture," in *Third International Conference on Intelligent Systems Modelling and Simulation*, 2012.

[4] W. A. Abdul and Parmanand, "Energy Efficieny in routing protocol and data collection approaches for WSN :A Survey," in *International Conference on Computing Communication and Automation(ICCCA)*, 2015.

[5] A. Samarth and S. Sumedha, "Data Aggregation Techniques to Remove Redundancy in Wireless Sensor Networks: Brief Overview.," *International Journal of Advance Foundation and Research in Computer*, vol. 1, no. 12, p. 2348 – 4853, 2014.

[6] R. Marius and P. Sever, "A WBAN-ECG Approach for Real-time Long-term Monitoring," in *The 8th International Symposium On Advanced Topics In Electrical Engineering*, Bucharest, Romania;, 2013.

[7] S. J. Sneha and D. P, "BSNs:A Special Approach to Monitor Heart Rate," in *International Journal of Latest Technology in Engineering, Management & Applied Science*, Mysore, 2014.

[8] L. Benoıˆt, B. Bart and M. Ingrid, "A survey on wireless body area networks," *Wireless Networks”*, vol. 17, no. 1, pp. 1-18, 2011.

[9] A. Sriyanjana, C. Sankhayan and C. Samiran, "A new routing Protocol For Wban to Enhance Energy Consumption And Network Lifetime," in *Proceedings of the 17th International Conference on Distributed Computing and Networking ACM*, 2016.

[10] T. T. Le and M. Sangman, "Interference Mitigation Schemes for Wireless Body Area Sensor Networks: A Comparative Survey," *Sensors,* vol. 15, no. 6, pp. 13805-13838, 2015.

[11] A. Ahmed and S. Arif, "Deployment Of Hash Function To Enhance Message Integrity In Wireless Body Area Network (Wban)," *International Journal of Communications, Network and System Sciences,* vol. 9, no. 12, pp. 613-621, 2016.

[12] K. Brad and H. T. Kung, "GPSR: Greedy Perimeter Stateless Routing for Wireless," in *International Conference on Mobile Computing and Networking*, USA, 2000.

[13] Q. Muhannad and J. Yaser, "Cloudlet-based Efficient Data Collection in Wireless Body Area Networks," *Simulation Modelling Practice and Theory*, p. 57–71, 2015.

[14] C. L. Kevin, L. Uichin and G. Mario, "Survey of Routing Protocols in Vehicular Ad Hoc Networks," in *Advances in Vehicular Ad-Hoc Networks: Developments and Challenges*, USA, 2010.

[15] T. Qiuling, T. Guoshun, W. Xi, S. Jiahao and H. Yulong, "An Energy-efficient Scheme for Data Collection in wireless sensor networks," in *Wireless and Optical Communication Conference*, 2016.

[16] W. Yufeng, C. T. Chiu and M. Ningfang, "Using Elasticity to Improve Inline Data Deduplication Storage Systems.," in *Cloud computing (CLOUD)*, 2014.

[17] Y. Yitao, Q. Xiaolin, S. Guozi, X. Yong, Y. Zhongxue and Z. Zhiyue, "Data deduplication in Wireless Multimedia Monitoring Network,"

International Journal of Distributed Sensor Networks, vol. 2013, p. 7, 2013.

[18] Y. Y. Kassaye, M. Antonis and G. B. Johan, "Secure and scalable deduplication of horizontally partitioned health data for privacy-preserving distributed statistical computation," *Medical informatics and decision making,* no. 17:1, 2017.

[19] G. Lu, Y. Jin and D. H. Du, "Frequency based chunking for data de-duplication," in *International Symposium on Modeling, Analysis & Simulation of Computer and Telecommunication Systems (MASCOTS)*, 2010.

[20] R. N. Widodo, H. Lim and M. Atiquzzaman, "A new content-defined chunking algorithm for data deduplication in cloud storage," *Future Generation Computer Systems,* vol. 71, pp. 145-156, 2017.

[21] C. Yu, C. Zhang, Y. Mao and F. Li, "Leap-based content defined chunking—theory and implementation," in *31st Symposium of Mass Storage Systems and Technologies IEEE (MSST)*, 2015.

[22] K. Eshghi and H. K. Tang, "A framework for analyzing and improving content-based chunking algorithms," in *Hewlett-Packard Labs Technical Report TR*, 2005.

[23] Y. Zhang, H. Jiang, D. Feng, W. Xia, M. Fu, F. Huang and Y. Zhou, "An asymmetric extremum content defined chunking algorithm for fast and bandwidth-efficient data deduplication," in *IEEE Conference on Computer Communications (INFOCOM)*, 2015.

[24] W. Xia, Y. Zhou, H. Jiang, D. Feng, Y. Hua, Y. Hu and Y. Zhang, "FastCDC: a Fast and Efficient Content-Defined Chunking Approach for Data Deduplication," in *USENIX Annual Technical Conference,* 2016.

[25] R. N. Widodo, H. Lim and M. Atiquzzaman, "SDM: Smart deduplication for mobile cloud storage," *Future Generation Computer Systems,* vol. 70, pp. 64-73, 2017.

[26] B. Romański, Ł. Heldt, W. Kilian, K. Lichota and C. Dubnicki, "Anchor-driven subchunk deduplication", Proceedings of t," in *4th ACM Annual International Conference on Systems and Storage*, 2011.

[27] E. Kruus, C. Ungureanu and C. Dubnicki, "Bimodal Content Defined Chunking for Backup Streams," in *In Fast*, 2010.

[28] D. R. Bobbarjung, S. Jagannathan and C. Dubnicki, "Improving duplicate elimination in storage systems," *ACM Transactions on Storage,* vol. 2, no. 4, pp. 424-448, 2006.

[29] C. Dubnicki, L. Gryz, L. Heldt, M. Kaczmarczyk, W. Kilian, P. Strzelczak and M. Welnicki, "HYDRAstor: A scalable secondary storage," in *FAST*, 2009.

[30] R. Bartłomiej, Ł. Heldt and K. Wojciech, "Anchor-Driven Subchunk Deduplication," in *Annual International Conference on Systems and Storage,* 2011.

[31] Ansari, R. Abdul and C. Sunghyun, "Humanbody: The future communication channel for WBAN," *InInternational Journal of Computer Applications (0975 – 8887)Volume 142 – No.11,* pp. 1-3, 2014.

[32] N. Bjørner, A. Blass and Y. Gurevich, "Content-dependent chunking for differential compression, the local maximum approach," *Journal of Computer and System Sciences,* vol. 76, no. 3-4, pp. 154-203, 2010.

[33] N. W. Ryan, L. Hyotaek and A. Mohammed, "A new content-defined chunking algorithm for data deduplication in cloud storage," *Future Generation Computer Systems,* vol. 71, pp. 145-156, 2017.

[34] T. Zhang, F. Damerau and D. Johnson, "Text chunking based on a generalization of winnow," *Journal of Machine Learning Research,* pp. 615-637, 2002.