Hybrid Data Compression System In Smart E-Health Gateway For Medical Monitoring Applications

Mezui Eya'a Guy Lysmos[#1], Dr.Mostafa Hanoune[*2]

[#]PhD Student, Mezui Eya'a Guy Lysmos, Dept. of Computer Science, University Hassan II of Casablanca

#Professor, Mostafa Hanoune, Dept. of Computer Science, University Hassan II of Casablanca

Faculty of the Sciences Ben M'Sik Av Driss El Harti B.P 7955, Sidi Othmane, Casablanca, Morocco

Abstract

161 millions healthcare devices will be connected by 2020 worldwide [1], which represents a real big data challenge, to store and process all medical data with the necessary historical depth. This is why we operate strategically in this article, one of the main services of the smart gateway, which is data compression to implement a hybrid data compression system for medical surveillance applications. It is by considering many challenges, both in terms of energy efficiency and security, which we effectively deal with the flow of data in real time [2].

The successful implementation of this hybrid data compression system in our smart gateway not only reduces the space required for data storage, but also accelerates lossless data transfer over the network and to the disk. In addition, the system has a model for automatic and intelligent recognition and classification of data sources according to the data structure and the original format.

However, a use case of lossless data compression algorithms was applied to assess the performance of the hybrid system. In addition, a comparative study of metric parameters for data compression was carried out to highlight the advantages and limitations of this system. Our design demonstrates an advanced IoT data processing system for medical and health monitoring, based on an improved global approach to advanced data compression methods [3].

Keywords — *Internet of things, healthcare, data processing, data compression, Smart Gateway, Lossless algorithms.*

I. INTRODUCTION

As the medical field advances dramatically in the digital age, healthcare companies, medical facilities and home patients generate large amounts of heterogeneous medical data exponentially each year. This data is either transmitted in medical applications, Or stored in the cloud or locally. Which is becoming a real problem for the network bandwidth which is always limited and also for the storage space which becomes insufficient. The majority of traditional and modern compression techniques are based on the following basic data compression system [4].

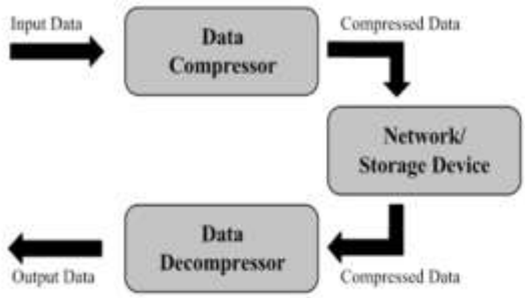


Fig 1: Components of Data Compression System

The Fig1, shows that ***Data Compressor*** compresses incoming data, which is a lossy or lossless compression algorithm. The compressed data is then shared across the network or stored in compressed format. Later, this data will be transcribed in its original format, thanks to ***Data Decompressor*** which is a data decompression agent on exit or when used.

The main contribution of the article is based on the implementation of a hybrid data compression system in the smart gateway. This system offers a heterogeneous layer of data processing between the data sources and the services that use them.

The rest of the article is organized as follows: Section 2 illustrates the sources of inspiration for this work and the cutting-edge compression algorithms we have taken as references; Section 3 presents the smart gateway; Section 4 presents and provides a description of the hybrid compression system, illustrating the different modules, their functionalities and their interactions. Section 5 reports on the results obtained in our evaluation campaign and gives an overview of the advantages and limits of the performance of the hybrid system on the results of real scenarios; and finally, section 6 that concludes the document with some future remarks.

II. GENERAL STUDY

OUR WORK WOULD NOT HAVE BEEN POSSIBLE WITHOUT THE FUNDAMENTAL CONTRIBUTIONS OF:

A. *Related works*

R.Saranya, M.Pavithra, R.Saranya, R.Sivalakshmi and N.Sridevi in “secure and safety health care monitoring system based on IoT” use monitoring applications. Network sensor devices either worn on the body or embedded in living environments, making possible of gathering rich information to evaluate physical and mental health condition of the patient [5].

Tuan Nguyen gia in "Smart e-Health Gateway: Bringing Intelligence to Internet-of-Things Based Ubiquitous Healthcare Systems", implemented a smart gateway under the name of UT-GATE to offer local data storage services and real-time data processing. One of the treatments is based on the use of two compression methods. His first method is to compress the data collected from the gateway nodes, and then send this data directly to the remote server. The second method collects the data in the form of nodes and saves the compressed data in the files that will be sent to the remote server [3].

Prasenjit Maiti in "Sensors Data Collection Architecture in the Internet of Mobile Things as a Service (IoMTaaS) Platform", presents an architecture that allows good buffer management, sensor message order, data alignment for multiple time data streams and data storage [6].

Stefan Karlsson and Erik Hansson in "Lossless Message Compression", presented a literal study on lossless compression using the compilation of the algorithms LZO, LZFX, LZW, LZMA, bzip2 and LZ4 in LINX as an additional layer to support the message without loss compression. The test was to send messages with real telecom data between two nodes on a dedicated network, with different network configurations and large messages [7].

Jeehong Kim and Jundong Cho propose in “Hardware-accelerated Fast Lossless Compression Based on LZ4 Algorithm” a hardware architecture with an improved compression rate and speed. The proposed architecture can achieve a compression rate up to 3.84 Gbps and a compression rate up to 4. This makes real-time data processing profitable and reduces the latency during the data compression process [8].

All these works are different in a few points but all converge towards a single objective, to set up a good data flow processing system.

B. *Appropriated algorithms*

In this part, we describe some suitable algorithms for our implementation. These algorithms were the most used in related work.

LZO: is an algorithm developed by M. Oberhumer, which is designed for fast compression and decompression. It uses LZ77 with a small hash table to perform searches. Compression requires 64 KB of memory during decompression requires no additional memory [9].

Bzip2: is a compression program based on the Burrows Wheeler algorithm. It generally compresses files between 10% and 15% of the best available techniques, while being approximately twice as fast in compression and six times faster in decompression. It compresses files using the forward displacement transformation, length encoding and Huffman coding [10].

LZ4: is a very fast compression algorithm based on LZ77.Data compression with LZ4 is very fast but data decompression is even faster and can reach speeds of up to 1 GB / s, especially for binary data [11].

III. PRESENTATION OF THE SMART GATEWAY

A. *Architectural*

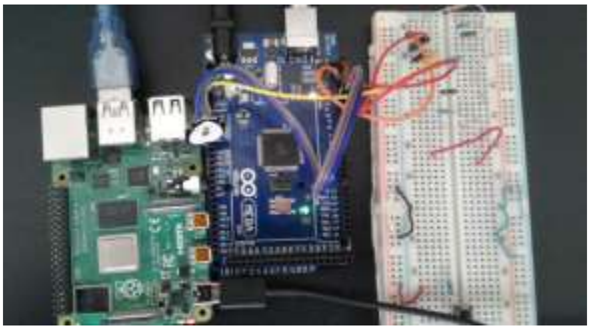
The smart gateway of Tuan Nguyen gia is the most similar to our gateway, because it is based on the same objective which is to share the flow of health information between patients, hospitals and health professionals to improve the quality of care of patients, from their homes to medico-social and health accommodation structures [3].

Our intelligent gateway differs from that of Tuan Nguyen gia on local data processing means specific to the architecture of the data compression system and on software and hardware components. In addition, our smart gateway provides an efficient recognition and classification system for data sources relative to their original format. See the block diagram below.

Fig 2: Architecture of the smart gateway for health

B. Software and hardware

The platform used in our project is the Raspberry Pi 4 model B which is the most accessible for innovators in the on-board space. It sports a faster system on a chip with a processor that uses the Cortex-A72 architecture (ARMv8 64-bit quad-core at 1.5 GHz) and has 4 GB of RAM. The gateway requires more or at least a kernel operating system to perform its operations.
And we use Linux also used by Tuan Nguyen gia in this work for reasons of its efficiency and its customization of factors.
The Linux kernel comes with basic settings called IPtables and FreeBSD. These are simple security mechanisms offering robust security depending on the configuration of the cryptographic module and firewall.



IV. SMART GATEWAY HYBRID DATA COMPRESSION SYSTEM

The intelligent gateway has an efficient hybrid data compression architecture with regard to data processing and the choice of data formats to compress. See the following diagram.

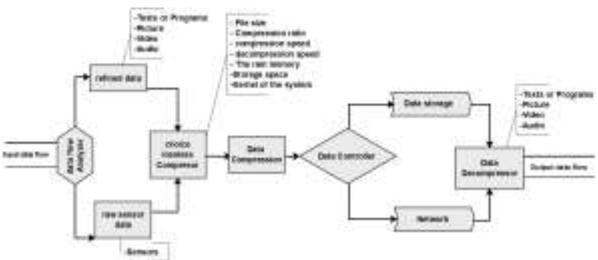


Fig 3: Architecture of the hybrid data compression system of the intelligent gateway for health

First, the original data flow is subject to recognition verification by a data flow analyzer module or system. The latter is an agent who analyzes the data and their format by in-depth research on the basic set of existing models (**refined data** and raw **sensor data**).

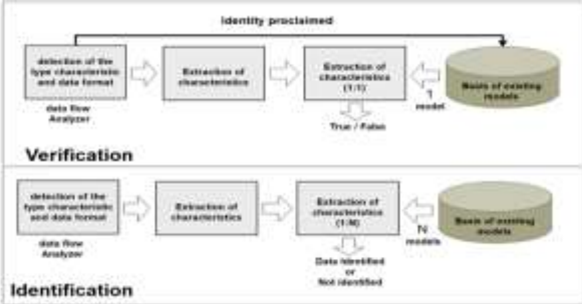


Fig 4: Functional diagram of the recognition, verification and identification of data.

Second, the identified data is compressed by a compressor specific to the types and formats of data according to the algorithmic (compression speed and compression rate) and hardware (RAM and storage space) performance criteria. Finally, thirdly, dedicated control agents for each functionality according to the following diagram manage the compressed data.

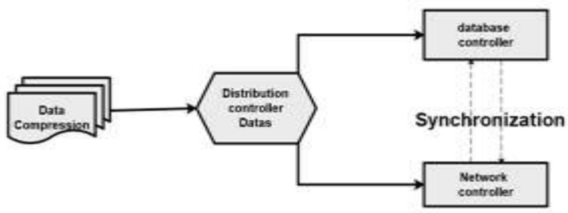


Fig 5: Functional diagram of the distribution of compressed data

Fig.5 shows that the first controller agent distributes the data in relation to the functional need of the two respective agent **database Controller** and **Network Controller**.

-For **database Controller**, it allows not only to monitor the availability of storage space, but also to update the database. The database is updated either by automatically deleting the old compressed files, thanks to the history of these files, or by adding the new files already compressed.

-For **Network Controller**, it is responsible for monitoring the network bandwidth and establishing good communication between the remote server and the smart gateway, thanks to notifications of major network performance indicators (latency, size of bandwidth and packet loss).

V. RESULTS

This section focuses on the evaluation of the hybrid architecture solution proposed in section 4. This evaluation is based on the comparison of the different metrics of the compression algorithms. The section is divided into three parts:

Part A, provides definitions of metric parameters, as well as some additional concepts for evaluation; Part B presents the tables and a graphic representation of the performance results of our hybrid compression system according to the metric parameters of the algorithms. Finally part C we will make a comparative deduction relating to the results of the tables to define the advantages and limits of the hybrid data compression system.

Regarding the datasets used in our assessment, we considered two different databases:
The first dataset is obtained from open data source TCIA. It is a service that identifies and hosts a large medical archive of cancer available for public download. The data are organized into "collections". The data are linked to images of patient results as well as details of treatment [12]. And the second set of data is obtained from a real deployment of the data from the sensors. We measured the readings from the biomedical sensors (Heart rate, body temperature and blood pressure) used by Jie Wan and Munassar AAH Al-awlaqi in "Wearable IoT enabled real-time health monitoring system" [13]. This dataset has been designated DSH (data Sensors Health).

A. *Definition of metric evaluation parameters*

The metrics used to evaluate the performance of the hybrid data compression system are: compression ratio, processing time and compression gain [14].
P. Kulalvaimozhi, M. Germanus Alex and S. John Peter also used metric performance to evaluate a study of image compression techniques in an agricultural field [15]. Let us designate Q for the compression quotient, T for percentage compression ratio and G for percentage compression gain.

$$Q=\frac{Initial\ Sizec}{Final\ Size}; \qquad T=\frac{1}{Q}\ \%; \qquad G=1-T=\frac{Initial\ Size-Final\ Size}{Initial\ Size}\%$$

<u>Processing time</u>**:** this is the time that the original data undergoes processing before being stored and sent to the remote server. Real-time processing is a key issue in online health applications. So latency during the data compression process should be considered.

B. *Presentation of tables and graphical representation of evaluation results.*

The results of the tables and the graph show that all the algorithms work well. We tested eleven files on three lossless compression algorithms. The first four are images of CT fused pulmonary pathology, extracted from the TCIA database [16]; the next four are the clinical data text files (CT COLONOGRAPHY, ADENOCARCINOME,ADENOCARCINOME CANALAIRE and GLIOBLASTOME, MULTIFORME) also extracted from the TCIA database and the last three are the data received in real time from the sensors.

We used the second Tuan Nguyen gia compression method, which consists of collecting the data in the form of nodes and saving the compressed data in the files which will be sent to the remote server. The choice of this method is due to the latency requirements during the real-time data compression process. According to IEEE 1073, the maximum latency for medical devices must be less than 500 MS [17].

Table I: Compression results for the LZO algorithm

Original Information			LZO			
Data Base	File Name	File Size (bytes)	Compressed File size	Compression Ratio (%)	Compression gain (%)	Processing Time(ms)
TCIA	CT01.tiff	542108352	308095590	56.8	43.2	3636
	CT02.tiff	472080576	285127982	60.4	39.6	3088
	CT03.tiff	377217984	204710601	54.3	45.7	2383
	CT04.tiff	563675328	336734861	59.7	40.3	3947
	CM.txt	108083	15753	14.6	85.4	10
	CTC.txt	905782	76999	8.5	91.5	50
	GBM.txt	165309	24586	14.9	85.1	30
	LUAD.txt	404609	59461	14.7	85.3	65
DSH	Sensor1.txt	43933	5800	13.2	86.8	33
	Sensor2.txt	45111	5962	13.2	86.8	35
	Sensor3.tx	38947	5710	14.6	85.4	27

Table II: Compression results for the LZ4 algorithm

Original Information			LZ4			
Data Base	File Name	File Size (bytes)	Compressed File size	Compression Ratio (%)	Compression gain (%)	Processing Time(ms)
TCIA	CT01.tiff	542108352	301539095	55.6	54.4	3857
	CT02.tiff	472080576	278938187	59.1	40.9	2473
	CT03.tiff	377217984	201896009	53.5	56.5	1768
	CT04.tiff	563675328	332051850	58.9	51.1	2678
	CM.txt	108083	14533	13.4	86.6	7
	CTC.txt	905782	68458	7.6	93.4	20
	GBM.txt	165309	21711	13.1	86.9	8
	LUAD.txt	404609	52050	12.9	87.1	26
DSH	Sensor1.txt	43933	6416	14.6	85.4	11
	Sensor2.txt	45111	6285	13.9	86.1	13
	Sensor3.txt	38947	6374	16.3	83.7	10

Table III: Compression results for the Bzip2 algorithm

Original Information			Bzip2			
Data Base	File Name	File Size (bytes)	Compressed File size	Compression Ratio (%)	Compression gain (%)	Processing Time(ms)
TCIA	CT01.tiff	542108352	197977507	36.5	63.5	51907
	CT02.tiff	472080576	192957939	40.8	59.2	43770
	CT03.tiff	377217984	136900717	36.2	63.8	25492
	CT04.tiff	563675328	219446276	38.9	61.1	59850
	CM.txt	108083	5820	5.3	94.7	19
	CTC.txt	905782	22945	2.5	97.5	123
	GBM.txt	165309	7935	4.8	95.2	37
	LUAD.txt	404609	16472	4.0	96.0	149
DSH	Sensor1.txt	43933	3718	8.5	91.5	48
	Sensor2.txt	45111	3860	8.6	91.4	60
	Sensor3.txt	38947	3618	9.3	90.7	33

C. Comparative study on the results of the tables

Table III shows that the processing time is relatively long for the Bzip2 compression algorithm. This presents a slight concern for the latency of real time data processing for the sensors, when a packet loss occurs in the network. Because the network controller must send a notification message to the database controller for a new packet dispatch.

However, this algorithm has a very big advantage on the data compression rate, which is beneficial for the optimization of the storage space estimated at more than 50% according to the compression gain.

In addition, the compression rate does not depend on the file size but on the structure and format of the file. We can see it in the CT2.tif file, the source code of the pro-gram, which contains the highest number of repeated elements, which results in a compression ratio between 40.8% and 60.4%.

Unlike Tables III, Tables I and II present a satisfactory data processing time com-pared to the data provided by the biomedical devices in real time. However, their compression ratio is not satisfactory enough because the objective is also to optimize the storage space. Therefore, we deduce that the LZO and LZ4 algorithms are best suited for processing real time data from biomedical sensors, for reasons of data dictionary efficiency. And the Bzip2 algorithm for the rest of the data coming from other data sources or external data generator, for reasons of better compression rate.

We also can see through the graph below, which highlights the evolution of data compression according to their original size.

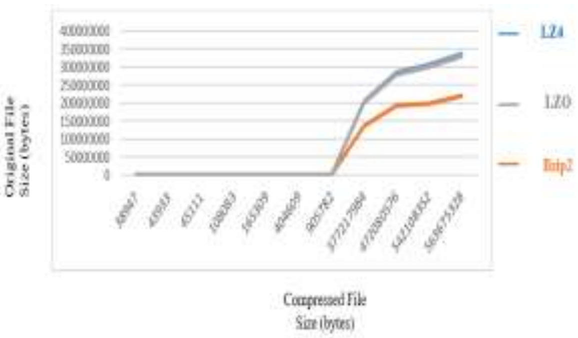


Fig 6: Comparative diagram of the performance of compression algorithms.

Advantages and limitations

Advantages
Support for all kinds of source file formats (origin) as well as their data structure.
Has an advanced system for recognizing data flows and data formats, thanks to the control modules and basic models recorded at each verification.
Ensures the integrity of data from the gateway to the remote server through the network controller, which notifies each transmission, the status of packets sent according to major indicators of network performance.
Optimization of the storage space to more than 50% by the controller agent, which regulates the storage memory through the logging of the backed up files.

Limitations
Absence of techniques for reducing the redundant data flow to be processed in the intelligent gateway. This can compromise not only the memory of storage devices, but also the processing time and network saturation.
Requires a lot of RAM memory when there is a large amount of data flow to be processed in real time.
Absence of privilege dedicated to the types of source data to be processed, in the event of simultaneous data processing.

VI. CONCLUSIONS

The implementation of the hybrid data compression system in the smart gateway for medical applications is a major advantage for real-time data processing. The main concept that guides the design of our solution is derived from the technique of recognizing the data source and their original formats. This technique initially optimizes the storage space thanks to a control module, which refers to an existing basic model.

Secondly, automate the intelligent choice of the best data compression algorithm according to metric criteria (Compression rate and processing time compared to the requirements of IEEE 1073 real-time data latency). Therefore, LZ4 and LZO were chosen for their compression speed capacity for bio-medical sensor data and Bzip2 for the rest of the data sources thanks to its compression rate efficiency and its performance on storage space.

Our future research will focus on the use of this hybrid data compression system, to develop compression mechanisms based on real data flows and to use classification information to optimize communication in the constrained part of the network.

VII. REFERENCES

[1] https://experiences.microsoft.fr/business/intelligence-artificielle-ia-business/sante-connectee-chiffres/.

[2] Capteurs pour la télésurveillance médicale. Capteurs, algorithmes et réseaux IRBM, Volume 30, Issue 3, June 2009, Pages 93-103.

[3] Smart e-Health Gateway: Bringing Intelligence to Internet-of-Things Based Ubiquitous Healthcare Systems DOI:10.1109/CCNC.2015.7158084 Conference: 12th Annual IEEE Consumer. Communications and Networking Conference (CCNC), At USA, Volume: 826-834.

[4] https://www.semanticscholar.org/paper/Lossless-Message-Compression-Bachelor-Thesis-in-Karlsson-Hansson/6548c6f508296218e7edb28f0263e5a3e7db5769 .

[5] Secure and safety health care monitoring system based on IoT. SSRG International Journal of Computer Science and Engineering (SSRG – IJCSE) – Special Issue ICFTESH Mar 2019.

[6] Sensors Data Collection Architecture in the Internet of Mobile Things as a Service (IoMTaaS) Platform. DOI: 10.1109/I-SMAC.2017.8058245. Conference: IEEE International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud.

[7] Modern Lossless Compression Techniques: Review, Comparison and Analysis. Conference Paper · February 2017 DOI: 10.1109/ICECCT.2017.8117850.

[8] Hardware-accelerated Fast Lossless Compression Based on LZ4 Algorithm. ICDSP 2019: Proceedings of the 2019 3rd International Conference on Digital Signal Processing February 2019 Pages 65–68.

[9] Markus Oberhumer. LZO Data Compression Library http://www.oberhumer.com/opensource/lzo/.

[10] Julian Seward. A program and library for data compression. http://www.bzip.org.

[11] Yann Collet. Real Time Data Compression-LZ4 http://fastcompression.blogspot.se/p/lz4.html.

[12] TCIA Collections: https://www.cancerimagingarchive.net/

[13] Wearable IoT enabled real-time health monitoring system Jie Wan, Munassar A. A. H. Al-awlaqi, Ming Song Li, Michael O'Grady, Xiang GU, Jin Wang & Ning Cao EURASIP Journal on Wireless Communications and Networking volume 2018, Article number: 298 (2018).

[14] S.R. Kodituwakku et al. / Indian Journal of Computer Science and Engineering.

[15] A Survey of Image Compression Techniques in an Agriculture Field. SSRG International Journal of Computer Science and Engineering (SSRG – IJCSE) – Special Issue ICFTESH Feb 2019.

[16] CPTAC-CM: https://wiki.cancerimagingarchive.net/display/Public/CPTAC-CM and https://app.box.com/s/gplxy09gr8adu4ay31eaqqxikjd4mh3p/folder/93339421102.

[17] IEEE Standard for Medical Device Communication, Overview and Framework. In ISO/IEEE 11073 Committee, 1996.