IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-7, Issue-1, Jan 2019 E-ISSN: 2347-2693

Data Compression Techniques in Wireless Sensor Network:A Survey

N.S. Singh[1*], H.J. Singh[2]

[1]Research Scholar, Department of *Computer* Science(Desh Bhagat University),Mandi Gobindgarh,India
[2]Assistant Professor, Department of Computer Science, Desh Bhagat University, Mandi Gobindgarh, India

[*]Corresponding Author: ns_kalra@yahoo.co.in , jeetsinder@gmail.com

Available online at: www.ijcseonline.org

Accepted: 26/Jan/2019, Published:31/Jan/2019

Abstract— Wireless sensor networks (WSNs) are highly resource guarded in terms of power supply, memory capacity, communication bandwidth, and processor performance. Compression of sampling, sensor data, and communications can considerably improve the effectiveness of utilization of power supply, memory and bandwidth. Recently, there have been a large number of research proposals describing compression algorithms for WSNs. These research proposals are diverse and involve various compression approaches.. In this paper, we take a step in that direction by presenting a survey of the literature in the area of compression and compression frameworks in WSNs. In addition, open research issues, challenges and future research directions are proposed.

Keywords— *WSN*, Compression, Data Compression

I. Introduction

Wireless Sensor networks are collection of distributed sensor nodes deployed under a particular location to monitor its environmental and physical locations such as pressure, temperature, humidity, sound and Sensor nodes come in different shapes or sizes. Sensor node is tiny device which contains the three important parts: a sensing subsystem for data acquirement i.e. data acquisition from the physical environment, a processing subsystem for processing storage of data and communication subsystem for transmission of data [1]. These subsystems also contain a power source which supplies energy to the device for working and a battery with inadequate energy budget[2,3]. Thus, it is quite inconvenient to recharge the battery, as they are organized in unrealizable surroundings. This makes the sensor nodes to be resource constrained. Wireless sensor networks (WSNs) have gained worldwide attention in recent years, particularly with the proliferation in Micro-Electro-Mechanical Systems (MEMS) Technology which has facilitated the development of smart sensors [4].They have limitations on power supply, bandwidth for communication, processing speed, and memory space

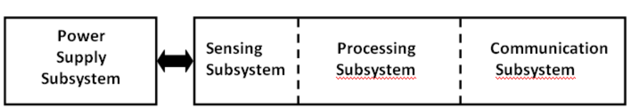


Figure 1 Wireless Sensor Network Architecture

WSN has the ability to serve various applications in situations which includes military Target tracking and investigation, natural disaster relief, biomedical health monitoring and hazardous environment exploration and seismic sensing and it contains many sensor nodes which are organized over a geographical area for observing physical occurrence like temperature, humidity, vibrations.. Thus it has given rise to many research works and studies which specifically concentrating on the maximizing the utilization of limited sensor resources. i.e. Energy. Thus it's important to adopt a number of approaches to reduce the power consumption in wireless sensor networks based on its architecture and these approaches mainly contain three important techniques i.e. duty cycling, data-driven approaches and mobility. Here we are focusing on data-driven approaches.

II. Data-Driven Approaches

Data-driven approaches describe the approach of reducing the number of sampled data while keeping the sensing accuracy within an acceptable range [5]. Two types of Data Driven approaches are:

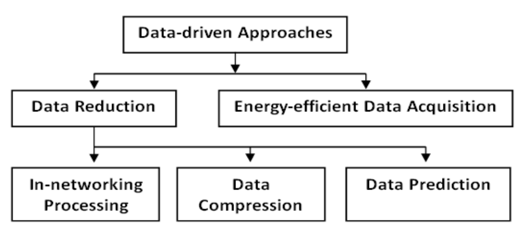


Figure 2 Data-driven approaches for WSNs.

Compression

Compression is a technique which is generally used to shrink data. Compression process saves execution time and memory space. Data compression is primarily used in communication as is aids devices to store or transmit the same amount of data but in less number of bits.

Data compression

Data compression is commonly referred to conversion of data in such a format that requires few bits usually formed to store and transmit the data easily and efficiently[6]. Compression is done either to reduce the volume of information in case of text, fax and images or reduce bandwidth in case of speech, audio and video.

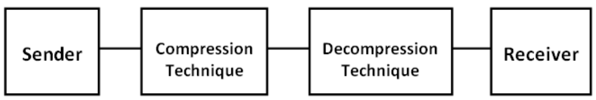


Figure 3 Compression Methods

III. FEATURES OF DATA COMPRESSION

A list of typical features of data compression for WSNs .

- Lossless vs. Lossy. Some compression algorithms are designed to support exact reconstruction of the original data after decompression (lossless). In other cases, the reconstructed data is only an approximation of the original (lossy). Use of a lossy algorithm may lead to loss of information, but generally ensures a higher compression ratio.
- Distortion vs. Accuracy. In the case of lossy compression, there is a trade-off between the data rate (R) achieved and the distortion (D) in the reconstructed signal. Mean Square Error (MSE) is a natural distortion metric. However, MSE can be misleading, since different types of distortion may have very different effects on the statistical inferences, which can be drawn after decompression. In addition, the energy consumption of communication should be taken into account.
- Data Aggregation. In some applications, only a summary of the sensor data is required. For example, statistical queries, such as MIN, AVG,MAX, allow for compact responses from the sensors. However, the original sample values cannot be reconstructed from the summarized representation. Aggregation requires in-network processing of sensor data but can greatly reduce communication overhead.
- Data Correlation. Since sensor nodes are normally deployed in close proximity, correlations between the sensed values at different nodes is often high (spatial correlation). Furthermore, since sensors observe events in a continuous manner, observed successive discrete signal samples often exhibit high correlation (temporal correlation).WSN compression algorithms typically exploit these correlations in order to improve the Compression ratio achieved.
- Symmetric vs. Asymmetric. In the case of symmetric algorithms, the computational complexity of compression and decompression are similar. In the asymmetric case, compression and decompression have different computational complexity. Traditional schemes tend to have higher computational complexity on the compression side. In contrast, in WSNs, it is desirable that compression, which is typically performed on the nodes, is low complexity and that decompression, which is typically performed at the sink, is high complexity.
- No adaptive vs. Adaptive. In no adaptive compression, the compression operations and parameters are fixed. This type of compression is suitable for stationary data, that is, when the statistics of the data do not change with time. In contrast, adaptive or dynamic compression methods monitor the raw data statistics and modify their operation and/or parameters in order to improve performance [7]. This approach is more complex but provides better performance for no stationary data.

IV. CLASSIFICATION OF COMPRESSION IN WIRELESS SENSOR NETWORK

In WSNs, the aim of compression is to reduce energy consumption. Sensing/sampling, computation, and communication are the operations mainly responsible for energy consumption in WSNs. Any technique that reduces one or more of these operations can be considered as compression. Compression in WSNs can be classified as follows.

Sampling Compression (SC). SC is the process of reducing the number of sensing/sampling operations while keeping network coverage (for spatially correlated sensors) and /or distortion loss within an acceptable margin. A number of research works [8,9] exploit spatially correlated data to reduce the sensing tasks. These works primarily focus on keeping the sensors in a sleep state, while a minimal number of sensors are kept active within a group. These are not the concern of this article. In contrast, CS [10,11] approaches perform the sampling-level compression by exploiting temporal data correlations at a sensor node.

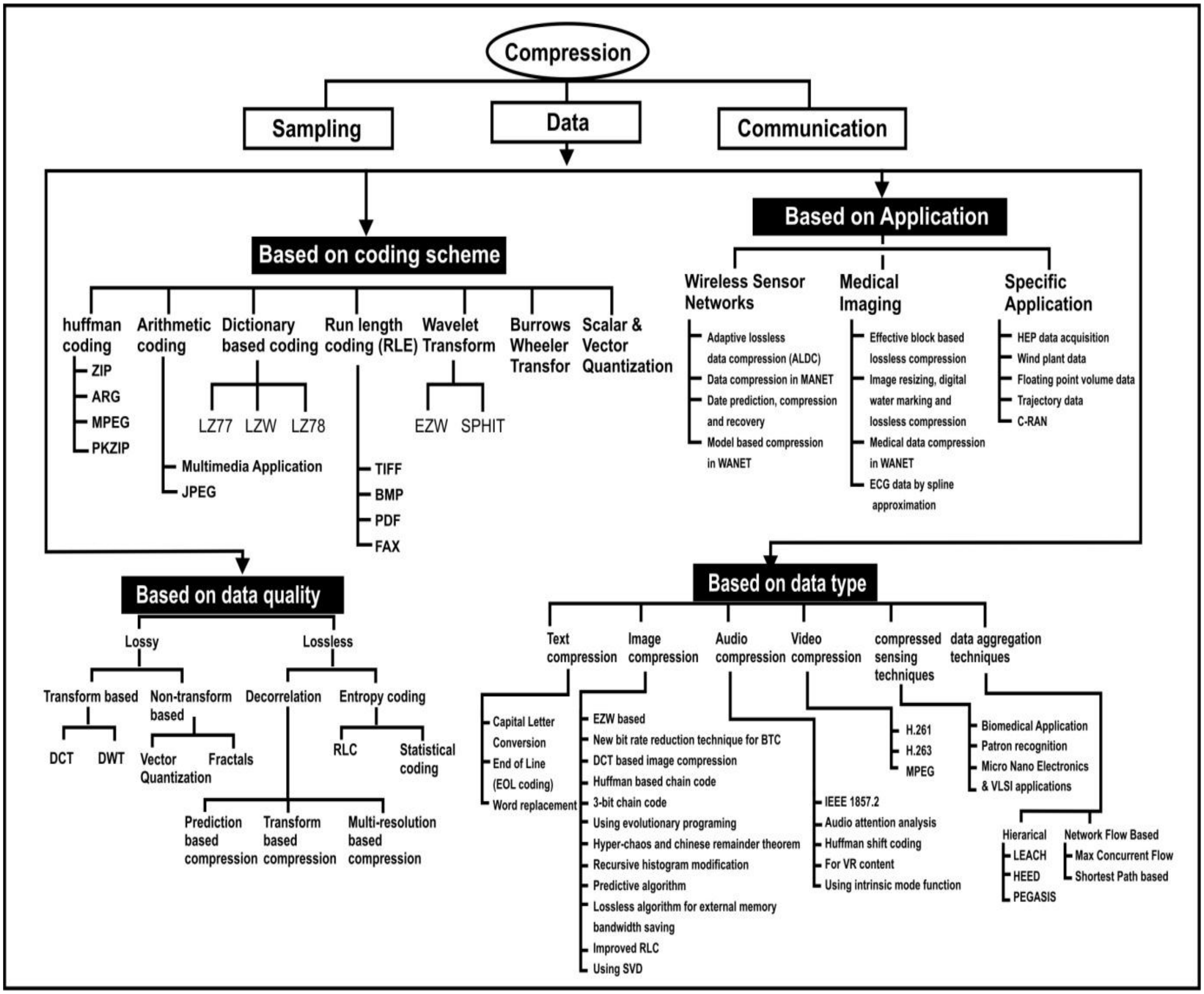


Figure 4 Classification of compression techniques

- Data Compression (DC). Data compression is the process of converting an input data stream (the source stream or the original raw data) into another data stream (the compressed stream) that has fewer bits. It can be viewed as the process of discovering structure that exists in the data and eliminating it by using more efficient encoding. All non random data has some structure, and this structure can be exploited to obtain a more compact representation of the data, that is, representation wherein no structure is noticeable. The terms redundancy and structure are used in the professional literature and are interchangeable [12,13].Most work on WSN compression(e.g., predictive coding, DSC, transform coding) supports data-level compression.
- Communication Compression (CC). Typically, this is the process of reducing the number of packet transmissions and receptions, hence reducing the radio on-time of transceivers a WSN. The longer the packet to be transmitted or received, the greater the radio on-time of transceivers [14, 15]. Hence reduced packet or data size (e.g., data compression) reduces radio on-time and reduces communication cost in WSNs. Aggregation, DCS, and predictive coding support communication-level compression.

➢ Data compression based on data quality

Data compression based on data quality is classified in two types- lossy compression and lossless compression.

- Lossless compression:- Lossless compression describes the compression in which original data can be recovered when the file is uncompressed. With lossless compression, data that was originally in the file remains after the file is uncompressed and the information is completely restored.

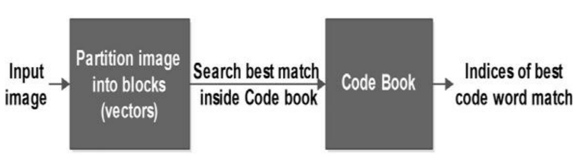
Transform coding: Transform coding is a type of data compression for data like audio signals or images. This technique is lossy, resulting in a lower quality of the original data. In transform coding, knowledge of the application is used to pick information to dispose of, thereby bringing down its band width. The left over information can be compressed by number of methods, the result may not be same as that of original input but close enough . [16]

Discrete Cosine Transform (DCT): DCT is a lossy compression technique which is broadly used for image and audio compression. DCTs are used to convert data in the abstract of series of cosine waves oscillating at different frequencies. There are very similar to Fourier transforms but DCT involves uses of cosine functions are much more efficient as fewer function are needed to approximate a signal. [16]

Discrete Wavelet Transform(DWT): The DWT is an implementation of the wavelet transform using a discrete set of the values scales and translations obeying some defined rules. In other words, this transform decomposes the signal into mutually orthogonal set of wavelets which is the main differences from the continuous wavelet transform or its implementation from the discrete time series sometimes called Discrete –time continuous wavelet transform (DTCWT). DWT is applied to the image block generated by the pre-processor. [16]

Nontransform based techniques: Nontransform based techniques are based on vector quantizer as quantization process in lossy compression techniques has two types of quantizers scalar and vector quantizers.

Vector Quantization (VQ). VQ encoding block diagram describes the image is partitioned into blocks of pixels and each block represented by vector x. This vector compared against code words in the code book at encoder side which will get the index of the best code word match. Therefore, index stored in fewer bits will be transmitted instead of code word which achieves more compression ratio.



- Lossy compression:- On the other hand, lossy compression reduces a file by permanently eliminating certain information, especially redundant information a certain amount of information loss that will not be detected by most users. Lossy compression is generally used for jpg, vedio and audio files [17].Lossless image compression techniques depend on two stage procedures [18]:

a. Decorrelation.
b. Entropy coding.

Decorrelation: In the first stage, decorrelation evacuates spatial repetition between pixels by one of lossless image compression techniques. These techniques are classified into three categories:

Prediction-based techniques
Transform based techniques and
Multi-resolution based techniques

Entropy coding: In the second stage, entropy coding which removes coding redundancy is based on Run Length Coding (RLC) and Statistical Coding [19]. Entropy is the average amount of information contain in messages received, where the message may be an event, sample or character drawn from a data stream. The therotical approach based on the concept of entrophy was introduced by Shannon. The amount of information S contained in a series of events $p_1,....,p_N$ should satisfy three requirements:

- S should be continuous in the p_i.
- if all the p_i are equally probably, so pi = 1/N, then H should be a monotonic increasing function of N.
- S should be additive.

$$S(t) = -\sum_{i=0}^{k} P_i \log_e(P_i)$$

Lossy Compression and Lossless Compression

Basis of comparison	Lossy Compression	Lossless Compression
Algorithms	Transform based , DCT, DWT, Non transform based vector Quantization fractal compression	RLW, LZW, Arithmetic encoding, Huffman encoding, Shannon Fano coding.
Used in	Images, audio and video.	Text or program, images and sound.

Data-holding capacity of the channel	More	Less as compared to lossy compression

- Based on coding Schemes

- Haffman coding:- A data compression technique which varies the length of the encoded symbol in ratio to its data content, that is the more frequently a symbol is used, the smaller the binary string used to represent it in the compressed form . This code can be properly decoded because they follow the prefix rule, which means that no code can be a prefix of another code, so the complete set of code can be represented as a binary tree which is known as Huffman tree. When a file is converted to the ZIP format, the data it contains is compressed. That is, the size of the compressed ZIP file is very much smaller than that of the original file.

- Arithmatic coding:-The fundamental problem of lossless compression is to decompose a data set (for example, a text file or an image) into a sequence of events, then to encode the events using as few bits as possible. The idea is to assign short codewords to more probable events and longer code words to less probable events. Whenever some events are more likely than others data is compressed. In Statistical coding techniques we use estimates of the probabilities of the events to assign the codewords. EgIn any case given a set of mutually distinct events el, e2, e3, . . . , e,, and an accurate assessment of the probability distribution P of the events, Shannon [17] proved that the the smallest possible likely number of bits needed to encode an event is the entropy of P, denoted by n

$$H(P) = \sum_{k=1}^{n} -p\{e_k\} \log_2 p\{e_k\}$$

where p{ek} is the probability that event ek occurs. A code outputs -log,p bits to encode an event whose probability of occurrence is p. Pure arithmetic codes with correct probabilities provide good compression.

- Dictionary based coding:- The compression technique [20][21] is used to proposed this theory by Ziv et al. To build the dictionary they use the substrings that appear many times in a message. As cyclic paths are not allowed in sensor networks, a node shows up at most once on a packet's path i.e., there is no repeated nodes on a path. As a result, these traditional dictionary based algorithms cannot be applied directly to provenance compression. The Lempel-Ziv algorithm (LZ) is a dictionary-based coding algorithm in lossless file compression. This is commonly used because of its usability in various file formats. It works for repeatedly occurring patterns and replaces them by a single symbol. It makes a dictionary of these patterns and the length of the dictionary is set to a exact value. This method is good for larger files and less helpful for smaller files. For smaller files, the size of the dictionary will be more than the original file. The two main versions of LZ were developed by Ziv and Lempel in two individual papers in 1977 and 1978, and they are named as LZ77 [22] and LZ78 [23]. These algorithms differ considerably in means of penetrating and finding matches. The LZ77 algorithm basically uses sliding window concept and searches for matches in a window within a predetermined distance back from the current position ZIP, and V.42 bits use LZ77. The LZ78 algorithm follows a more traditional approach of appending strings to the dictionary. LZW is an enhanced version of LZ77 and LZ78 which is developed by Terry Welch in 1984 [24]. The encoder makes an adaptive dictionary to describe the variable-length strings with no prior knowledge of the input. The decoder also constructs the similar dictionary as encoder based on the received code vigorously. The encoder store these symbols and maps it to one code. Typically, an LZW code is 12-bits length (4096 codes). GIF images, PNG images file formats use LZW coding.
- Run length Encoding

RLE is simple technique to compress digital data . It represent consecutive runs of same value in the data as the value followed by the count (or vice versa) .Run length is defined as number of consecutive equal values .the objective is to reduce the amount of data needed for storage transmit ion The Idea behind this algorithm is, If a data item d occurs n consecutive times in the input data we replace the n occurrences with the single pair nd. [25]

Run-Length Encoding (RLE) is a basic pression algorithm. It is very useful in case of repetitive and

slowly varying data items. This is basic compression algorithm on data that contains many such runs: for example, relatively simple graphic images such as grayscale images, icons..
It is not practical with files that don't have many runs as it could double the file size.
.For example, consider a screen Containing plain black text on a solid white background. It contains long runs of white pixels in the blank space, and fewort runs of black pixels within the text. Let us take a hypothetical single scan line, with W representing white B representing a black pixel :
WWWWWWWWWWWWWBWWWWWWWWW
WWWWBBBWWWWWWWWWWWWWWW
WWWWWWWWWWBWWWWWWWWWWWW
WWW
Apply the run-length encoding (RLE) data compression algorithm to the above hypothetical scan line, the encoded text is as follows:
13W1B12W3B24W1B15W Interpret this as twelve W's, one B, twelve W's, three B's,etc.
The run-length code represents the original 69 characters in only 20. The principle remains the same The graphical representation of the RLE algorithm can also be applied on temperature readings.

- Wavelet transform(BWT) is also known as block sorting compression which rearranges the character string into runs of identical characters. It uses two techniques to compress data: RLE & move-to-front transform. It compresses data easily in situations where the string consists of runs of repeated characters. The significant feature of BWT is the reversibility which is fully reversible and it does not require any extra bits. BWT is a "free" method to improve the efficiency of text compression algorithms, with some additional computation. It is s used in bzip2. A simpler form of lossless DC coding technique is RLE. It represents the sequence of symbols as runs and others are termed as non-runs. The run consists of data value and count instead of original run. It is effective for data with high redundancy. Embedded Zero tree wavelet (EZW) [26], Set Partitioning in Hierarchical Trees (SPIHT), 3D_SPIHT are three types of Wavelet transform which are applied on the image. EZW The embedded zero tree wavelet algorithm (EZW) produces embedded bit stream for image coding. This technique requires neither training nor pre-stored codebooks or tables and requires no preceding knowledge of the source image. The EZW algorithm is based on a discrete wavelet transform[27] or decomposition of hierarchical sub-band , prediction of absence of significant information , Entropy coded successive-approximation quantization and universal lossless data compression which is accomplished using adaptive Arithmetic coding. SPIHT is also an embedded wavelet coding algorithm. The SPIHT algorithm[28] is the next version of the EZW algorithm. The SPHT algorithm is different because it does not directly transmit the contents of the sets the pixel coordinates[29] but it send decisions which is made in every node of the trees that define the structure of the image. Because here only decisions are being transmitted, the pixel value is defined by in which points the decisions are taken and their outcomes, while the coordinates of the pixels are defined, which tree and what part of the tree, the decision is being made on. The advantage to this is that the decoder has an identical algorithm which is able to identify with each of the decisions and can create identical sets along with the encoder.
- Burrows–Wheeler transform (BWT, also called block-sorting compression):- It rearranges a character string into runs of similar characters. and tends to be easy to compress a string that has runs of repeated characters by techniques such as move-to-front transform and run-length encoding. The transformation is reversible, without any additional data information except the position of the first original character. The BWT is thus a "free" method of improving the efficiency of text compression algorithms, costing only some extra computation. When a character string is transformed by the BWT, the transformation permutes the order of the characters. If the original string had many substrings that occurred repeatedly, then the transformed string will have a number of places where a single character is repeated multiple times in a row. For example:

Input	SIX.MIXED.PIXIES.SIFT.SIXTY.PIXIE.DUST.BOXES
Output	TEXYDST.E.IXIXIXXSSMPPS.B..E.S.EUSFXDIIOIIIT

The output is easier to compress because it has many repeated characters. In this example the transformed string contains six runs of identical characters: XX, SS, PP, II, and III, which together make 13 out of the 44 characters.

- **Based on Applications**

Based on wireless sensor networks application Sensor nodes are usually battery-powered and thus how to conserve their energy is a primary concern in WSNs. In network data compression can help reduce the amount of sensing data that sensor nodes have to regularly report to the sink(s) and therefore significantly reduce their energy consumption. There is Five categories of data compression techniques, including string based compression, image-based compression, distributed source coding, compressed sensing, and data aggregation. ALDC Scheme performs compression losslessly using two code options. ALDC algorithm is efficient and simple, and is particularly suitable for resource-constrained wireless sensor nodes. ALDC compression scheme allows compression to dynamically adjust to a changing source and reduce the data amount for transmission which contributes to the energy saving. Algorithm can be used in monitoring systems that have different types of data and still provide satisfactory compression ratios. MANET, Data prediction, recovery and mobile based compression in WANET are examples of wsn application.

- Based on medical imaging: The new health care reform in China is causing an increase in the demand for PACS (picture archiving and communication systems). Medical equipment of digital imaging has been producing massive medical imaging in hospitals. This poses a serious challenge to the limited transmission bandwidth and storage capacity of PACS. So there is an urgent need for efficient medical imaging compression algorithm[24][25][26].A novel medical data compression (MDC) method is employed for the reduction in size of the patient’s medical data (MDP) [30]. The compression and decompression are done at Primary Health Care (PHC) and community care centre (CC). The compression is divided into two levels: Image transformation and encoding of coefficients. The compressed data is transmitted from PHC to CC center using WANET. Because of transformation, quantization and transmission errors, blocking artifacts, additive noises and multiplicative noise may present. To eliminate the noise during decompression, an adaptive edge-based fuzzy filtering technique is used. A Fuzzy based Route Selection (FRS) is also proposed to find the optimal path from PHC to CC. Fuzzy logic uses input parameters as energy consumption, residual energy and routing delay. The probability of selecting a routing path is the output parameter of fuzzy logic. The path with higher probability has more chance of becoming the routing path. Once a path is established, the compressed image is transmitted from PHC to CC center efficiently. The proposed method is compared with JPEG, JPEG2000, Turcza and Duplaga in terms of CR visual quality of medical data of patients and robustness against signal transmission errors. This method is used to compress endoscopic images, ECG images, magnetic resonance imaging, etc. [31] proposed a compression technique named Watermarked Reduction/Expansion Protocol integrated to TIFF format (WREPro.TIFF) and Watermarked Reduction/Expansion Protocol combined with JPEG.LS format (WREPRo.JLS) for medical images. In this approach, four techniques such as expansion technique, digital watermarking and lossless compression standards like JPEG-LS and TIFF format are employed. When compared with original JPEG, proposed method achieves better compression in terms of PSNR and SSIM. Computerized electrocardiogram (ECG), electroencephalogram (EEG), electromyographic (EMG) and magnetoencephalogram (MEG) processing systems are commonly employed in hospitals to record and process biomedical signals. These systems lead to the generation of larger size signal databases for successive investigation and comparisons. It also makes harder to transmit the biomedical information possible over telecommunication networks in real time. Biomedical signal compression methods are developed to reduce the storage space with no loss of clinically significant information, which is attained by the elimination of redundant data in the signal at a reasonable rate.

Based on data type

Data compression techniques in WSNs can be classified into following categories [32]:

- The text-based compression techniques view sensing data as a sequence of characters and then adopt the data compression schemes used to handle text data to compress these sensing data. Inherited from these text data compression schemes, the string-based compression techniques can also provide lossless compression. This compression technique includes capital letter conversion, EOL coding , word replacement, etc.

- The image-based compression techniques organize a WSN into a hierarchical architecture and then adopt some image compression schemes such as wavelet transformation to provide multiple resolutions of sensing data inside the network. Some minor features of sensing data may be lost due to the compression operations and thus the image-based compression technique support loss compression.
- Audio compression
  IEEE 1857.2 is latest standard for advanced audio coding (AAC) [33] is an efficient lossless audio compression technique basically designed to get better audio quality with high speed and optimized bandwidth. It involves a collection of tools to attain specific audio coding functions. In this method, encoding is done by Linear Prediction Coding (LPC) followed by preprocessing and entropy coding block. The performance is analyzed in both music and speech audio file. It produces better compression and faster rate of encoding and decoding process. Another audio compression method based on scalable variable bit rate encoding method is proposed in[34] (Hang et al., 2016) which is adaptable to bandwidth variations. An audio attention model is used to determine the attention of every audio sample which results in the determination of attention presence. A new method is proposed in [35] to produce original audio signals after the compression process which uses LZ compression over Intrinsic Mode Functions (IMF) produced by Empirical Mode Decomposition (EMD).
- Video compression

Generally, video files use more resources for statement, processing, and memory purposes. So, compression is required for storing video files, their process or transmit ion. H.261standard has proposed to transmit video at a rate of 64kbps and its multiples. The frames of H.261 can be classified into two types namely Intra coded frames (I-frames) and Predictedcoded (P-frames). The image is partitioned into a number of macro-blocks. Macro block has 16 _ 16 luminance blocks and 8 _ 8 chrominance blocks. The macro blocks are further encoded as intra or inter blocks. Spatial duplicity is utilized by DCT coding and temporal redundancy by motion compensation. H.263 contains motion compensation with half-pixel accuracy and bi-directional coded macro blocks. As H.263 is developed for low bit rate applications, these characteristics are not utilized in MPEG-1 and MPEG-2. Moving Pictures Experts Group (MPEG) is an ISO/IEC working group which develops compression techniques, representation of motion pictures and audio in international standards [36]. MPEG is a layered, DCT based video compression standards lead to VHS quality compressed video stream with a bit rate of nearly 1.5Mbps at a resolution of nearly 352 _ 240. MPEG video sequences comprised of different layers which can randomly access a video sequence and protects it against corrupted data. MPEG frames can be coded in three ways: I-frames, P-frames and Bi-directionally-predictive-coded or B-frames.

The compressed sensing techniques indicate that any sufficiently compressible data can be accurately recovered from a small number of nonadaptive, randomized linear projection samples. Thus, they can exploit compressibility without relying on any prior knowledge or assumption on sensing data. With the 1above observation, the compressed sensing techniques can provide lossless compression.

The data aggregation techniques

Data aggregation techniques select a subset of sensor nodes to collect and fuse the sensing data sent from their neighbouring nodes and then transmit the small sized aggregated data to the sink. Because the original sensing data cannot be derived from these aggregated data, the compression of the data aggregation techniques is thus unrecoverable. LEACH is the primary and most popular energy-efficient hierarchical clustering algorithm for WSNs which was proposed in order to reduce the power consumption. HEED [38][39] extends the basic scheme of LEACH by using residual energy and node degree or density as a metric for cluster selection to achieve power balancing. It operates in multi-hop networks, using an adaptive transmission power in the inter-clustering communication. PEGASIS [40] is an extension of the LEACH protocol, which forms chains from sensor nodes so that each node transmits and receives from a neighbour and only one node is selected from that chain to transmit to the base station (sink).

V. CONCLUSION AND FUTURE SCOPE

Development of effective compression algorithms is key to the improved utilization of the limited Resources of WSNs (energy, bandwidth, computational power). A large number of research proposals have addressed the problem. The proposals are diverse and involve various compression approaches. In this work, we have made an effort to put these works into perspective and to present a holistic view of the field. In doing this, we have provided a broad overview of existing approaches, the current state-of-the-art and presented a logical classification. Works are categorized as involving either aggregation, text-based compression, Distributed Source Coding, transform-based compression, Compressive Sensing and Predictive Coding. The approach adopted within each category has a number of variants which are presented accordingly. Although the presented approaches and frameworks address many issues associated with data compression in WSNs, some research questions remain relatively unexplored, such as support for and integration of QoS, scalability, reliability and security. There is significant scope for future work in one of these areas. Realizing the importance of QoS in WSNs, our future endeavors will focus on developing a compression framework, which integrates QoS-awareness for WSNs. Data compression in WSNs is a regular activity, integration of QoS awareness within it will ultimately contribute in developing a QoS-aware data gathering framework for WSNs. The diverse applications of WSNs demand support for a diverse set of QoS requirements.A single compression technique will not be optimal for all applications. Along with QoS awareness, a secondary objective will be to determine the best possible compression technique for a particular application taking into account the limited available resources.

REFERENCES

[1]. Giuseppe Anastasi, Marco Conti, Mario Di Francesco, Andrea Passarella, "Energy Conservation issues in WSN: A Survey ", IIT, National Research Council (NRC).

[2]. I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, "A survey on sensor networks", IEEE Communications Magazine 40 (8), pp. 104-112,2002.

[3] Rekha Rani, Rajan Manro, "Improved Watermarking Leach Protocol using node level Integrity and Confidentiality in WSN, international journal of computer science and engineering, Vol-6, Issue-11 Nov-2018 E-ISSN 2347-2693.

[4] Jennifer Yick, Biswanath Mukheerjee, and Dipak Ghosal,"Wireless Sensor Network Survey", Elsevier,Computer Networks 52,pp. 2292-2330, 2008.

[5]. Rumpa Mukherjee,Arindom Mukherjee ," A Survey on Different Approaches for Energy Conservation in Wireless Sensor Networks" ,International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970), Volume-3 Number-1 Issue-8 March-2013.

[6]. Emad M. Abdelmoghith, and Hussein T. Mouftah," A Data Mining Approach to Energy Efficiency in Wireless Sensor Networks", IEEE 24th International Symposium on Personal, Indoor and Mobile Radio Communications: Mobile and Wireless Networks, 2013.

[7] LI, X.-Y., WANG, Y., AND WANG, Y. 2010. Complexity of data collection, aggregation, and selection for wireless sensor networks. IEEE Trans. Comput.

[8] CARDEI, M., THAI, M. T., LI, Y., AND WU, W. 2005. Energy-efficient target coverage in wireless sensor networks. In Proceedings of the IEEE INFOCOM pp.1976–1984.

[9] SUBRAMANIAN, R. AND FEKRI, F. 2006. Sleep scheduling and lifetime maximization in sensor networks: Fundamental limits and optimal solutions. In Proceedings of the 5th International Conference on Information Processing in Sensor Networks. ACM, 218–225.

[10]. CAND`ES, E. AND WAKIN, M. 2008. An introduction to compressive sampling. IEEE Signal Process. Mag. 25, 3.

[11]. HAUPT, J., BAJWA, W., RABBAT, M., AND NOWAK, R. 2008. Compressed sensing for networked data. IEEE Signal Process. Mag. 25, 2.

[12]. SALOMON, D. 2007. Data Compression: The Complete Reference 3rd Ed. Springer-Verlag.

[13].SAYOOD, K. 2006. Introduction to Data Compression 2nd Ed. Morgan Kaufmann Publishers Inc.

[14] KIMURA, N. AND LATIFI, S. 2005. A survey on data compression in wireless sensor networks. In Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'05). IEEE Computer Society, 8–13.

[15]. BARR, K. C. AND ASANOVI´C, K. 2006. Energy-aware lossless data compression. ACM Trans. Comput. Syst. 24, 3, 250–291.

[16] Akyildiz IF, Melodia T, Chowdhury KR., "A survey on wireless multimedia sensor networks", Comput Netw 2007;51(4):921–60.

[17] C.E. Shannon, "A mathematical theory of communication," Bell Syst. Tech. J., vol. 27, pp. 398-403, July 1948.

[18] Yang M, Bourbakis N., "An overview of lossless digital image compression techniques", IEEE Circ Syst 2005;2:1099–102.

[19] samra HS. ''Image compression techniques''. Int J ComputTechnol 2012;2(2).

[20] J.Ziv and A.Lempel, "A Universal Algorithm for Data Compression," IEEE Trans. Inf. Theory, vol. 23, no. 3, pp. 337–343, 1977.

[21] J.Ziv and A.Lempel, "lz78.pdf." IEEE, pp. 530–536, 1978.

[22] J Ziv. A Lempel., "A Universal Algorithm for DC".IEEE Trans. Inf. Theory, 23(3): 337–343,1977.

[23]Ziv J, Lempel A. lz78.pdf. IEEE. 530–536,1978.

[24]Welch TA, "A technique for high-Performance DC", IEEE. 8–19,1984.

[25] D. A. Huffman, "A method for the construction of minimum redundancy codes," Proc.IRE, vol. 40, pp. 1098-1101, 1952.

[26] J. Tian and R.O. Wells, Jr. "A lossy image codec based on index coding" IEEE Data Compression Conference, DCC '96, page 456, 1996.

[27] J. M. Shapiro, "Embedded image coding using zero trees of wavelet coefficients" IEEE Transactions on Signal Processing, vol. 41, No. 12,pp. 3445-3462, 1993.

[28] Beong Jo Kim, Zixiang Xiong,William a. Pearlman," Low bit Rate, Scalable Video coding with 3D Set Partitioning in Hierarchical Trees" IEEE Trans. on Circuits and Systems for Video Technology, Vol. 10, No. 8, pp. 1374–1387, 2000.

[29] Kassim A Lee W S. "Color image coding using SPIHT with partially linked spatial orientation trees ". IEEE Trans on Circuit and System for Video Technology, Vol.2, No 2, 203-206, 2003.

[30] Dutta, T., "Medical data compression and transmission in wireless Ad Hoc networks", Sens. J., IEEE 15, 778–786,2015.

[31] Amri, H., Khalfallah, A., Gargouri, M., Nebhani, N., Lapayre, J.-C., Bouhlel, M.-S., 2017.Medical image compression approach

based on image resizing, digital watermarking and lossless compression. J. Signal Process. Syst. 87, 203–214.

[32] Data Compression Techniques in Wireless Sensor Networks PERVASIVE COMPUTING You-Chiun Wang Y.-C. Wang is with the Department of Computer Science, National Chiao-Tung University, Hsin-Chu, 30010, Taiwan. E-mail: wangyc@cs.nctu.edu.tw

[33] Auristin, F.N., 2016. Advanced audio compression for lossless audio coding using.

[34] Hang, B., Wang, Y., Kang, C., "A scalable variable bit rate audio codec based on audio attention analysis", Rev. Tec. Fac. Ing. Univ. Zulia 39, 114–120,2016.

[35] Jain, D., Tech, M., Engineering, C., Engineering, C., 2017. A Modified Method for Sound Compression Using Intrinsic Mode Functions 97–102.

[36] Watkinson, J., 2004. The MPEG Handbook: MPEG-1, MPEG-2, MPEG-4. Taylor & Francis.

[37] O. Younis and S. Fahmy, "Distributed Clustering in Adhoc Sensor Netwrok: A Hybrid, Energy -efficient Approach", September 2002.

[38] O. Younis and S. Fahmy, "Heed:A hybrid , Energy efficient, Distributed Clustering Approach for ad-hoc network", IEEE Transactions on mobile computing, vol.3, no.4, Oct-Dec.2004, pp. 366-369.

[39] Lindsey and C.S. Raghavendra, "PEGASIS: Power- efficient Gathering in sensor information System", Proceedings, IEEE Aerospace Conference, vol. 3, Big Sky, MT, Mar.2002, pp. 1125-1130.

[40] Yakubu Ajiji Makeri The role of Cyber Security and Human-Technology Centric for Digital Transformation, International Jounal of scientific research in computer science technology Vol-6, Issue-6, E-ISSN 2320-7639.

[41] A. Jenita Jebamalar**,** Efficiency of Data Mining Algorithms Used In Agnostic Data Analytics Insight Tools, IJSRNSC, vol. 6, Issue. 6, pp. 14-18, Dec-2018

Authors Profile

Mr. Narinder Singh had done Master of computer application from Maharishi dayanand univ*ersity*, *Rohtak* in 2004 and M.phil from choudhary Devi Lal university in 2008. He is currently pursuing Ph.D. and currently working as Assistant Professor in Department of Computer Sciences, Guru Nanak College, Budhlada since 2007. He is a member of computer society of India since 2010 and managing editer of international journol of information management and technology. He has published more than 10 research papers in national and international reputed journals and published three books. His main research work focuses on wireless sensor network,Network Security, Cloud Security and Privacy. He has 11 years of teaching experience and 5 years of Research Experience.

Dr. Harmaninder singh is an Associate Professor and has more than 12 years of academic experience. He has done MCA degree from Punjabi University Patiala, Punjab, India and received the P.hd Degree in Computer Science from Desh bhagat university, Mandi gobindgarh, India in 2008. More than 17 papers are published in different Journals and conference proceedings. His research interest is in the areas of Open Source Software and Algorithm Design and is presently working on it.