



EPiC Series in Computing

Volume 58, 2019, Pages 99–106

Proceedings of 34th International Conference on Computers and Their Applications



Weighted Record Sample for Underwater Seismic Monitoring Application

Hussain Albarakati¹, Reda Ammar² and Raafat Elfouly³

^{1, 2}, Computer Science and Engineering Department, University of Connecticut, Storrs, CT 06269

³Computer Science Department, Rhode Island College
Hussain.albarakati@uconn.edu, reda.ammar@uconn.edu, relfouly@ric.edu

Abstract

Underwater acoustic sensor networks have been developed as a new technology for real-time underwater applications, including seismic monitoring, disaster prevention, and oil well inspection. Unfortunately, this new technology is constrained to data sensing, large-volume transmission, and forwarding. As a result, the transmission of large volumes of data is costly in terms of both time and power. We thus focused our research activities on the development of embedded underwater computing systems. In this advanced technology, information extraction is performed underwater using data mining techniques or compression algorithms. We previously presented a new set of real-time underwater embedded system architectures that can manage multiple network configurations. In this study, we extend our research to develop information extraction for seismic monitoring underwater application to meet real-time constraints. The system performance is measured in terms of the minimum end-to-end delay and power consumption. The simulation results are presented to measure the performance of our architecture based on the information extraction algorithm.

Keywords: Underwater acoustic sensor networks, underwater embedded system, architecture, real-time constraints, information extraction, multipath.

1 Introduction

Underwater acoustic sensor networks (UWSNs) have many real-time applications, such as oil well inspection, fish classification, seismic monitoring, disaster prevention, and military applications (Heidemann, 2006) (Proakis, 2001) (Jindal, 2014). These applications demand a high bit rate and low power consumption. Therefore, real-time constraints are not achieved in UWSNs due to the limitations of acoustic signals, such as a low data rate, high propagation delay, limited battery life, and high error rate (Jindal, 2014). In a previous study, we presented real-time UWSN embedded architectures that optimize the end-to-end delay and the network's power consumption (Albarakati H. A., 2017)

(Albarakati H. R., December 2017). The motivation was to design dynamic architectures based on network characteristics (such as bit rate, gathering node capabilities, central computer [CC] processing capabilities, and ocean depth) for both homogenous and heterogeneous applications. Analytical models were constructed for our proposed architectures to enable the network designers to simultaneously measure the network up-front and evaluate the network capabilities for various applications. We could then choose the matching configuration that met the design constraints and/or objectives. To meet real-time constraints, we aimed to enhance the performance of the computing in three ways: (1) sensor node (SN) deployment, (2) data collection, and (3) information extraction. We previously proposed a heuristic algorithm and SN topologies to enhance data collection in the lowest layer of our architecture. This algorithm separates the sensors into clusters of sensors and then identifies the best location for master nodes (MNs) (one for each cluster) before it finally identifies the best location of the CC (Albarakati H. R., 2018).

In this paper, we highlight the development of the information extraction section. We present a sampling method based on weighted selection. First, we find the weight of each record and select the random sample from the data based on the weight of the record. We then extract the information from the sample. Here, we choose the seismic monitoring application because it is the most popular of real-time underwater applications. Simulations are used to verify the results and evaluate the performance of the architectures.

The remainder of the paper is structured as follows. Section 2 examines the related work. Section 3 defines and describes our developed architecture. In Section 4, we define and describe the weighted records sample algorithm. Section 5 presents and discusses our results. Section 6 summarizes our contributions and suggests directions for future work.

2 Related Work

One drawback of UWSNs is that they suffer from delay and power consumption (lifetime of network). In fact, replacement or recharging of the sensor node battery is impractical deep in the ocean (Heidemann, 2006). These limitations represent challenges to the deployment of efficient UWSNs. Currently, there are two methods to handle these challenges, mainly autonomous underwater vehicles (AUVs) (Zhuo, 2016) (Khan, 2015) and the multipath method (Magadza, 2015). In the previous study, we discussed more details about two methods in (Albarakati H. R., 2018). Here, we provide a summary about two methods.

First, the AUV collects data from the network in deep layer and sends them to the mothership. The AUV methods reduce the delay, but the AUV methods are impractical deep in the ocean (Petillo, 2010).

The multipath method deploys multiple relay nodes in various directions between the source and the destination. These nodes send the data from the source in the deep layer to its destination at the surface via multiple paths. The multiple paths method enhances the bottleneck in communication, but it does not minimize the delay or power consumption (lifetime of network) as well.

We therefore presented a new reconfigurable UWSN architecture that optimizes both the end-to-end delay and power consumption in the deep water. This new alternative method can be used for all underwater applications (i.e., homogenous and heterogeneous applications). Specifically, it avoids the problems discussed above and is a reasonable tool to meet real-time constraints.

In this study, we propose a new approach to extract information to meet real time constraints.

3 Architecture

In a previous study, we developed the multipath multiprocessor architecture exhibited in Figure 1, which consists of a central computer (CC), the sensor nodes (SNs), the master nodes (MNs) and the gateways (Gs) (Albarakati H. A., 2017) (Albarakati H. R., December 2017) (Albarakati H. R., 2018). The CC consists of VM hardware, a hypervisor, and hardware. It runs several VMs, each of which is designated to one MN and an application. Each cluster of nodes has an MN that retrieves the data from several SNs in its associated cluster and delivers it to the CC. The main agenda of the VMs is to analyze these data and obtain valuable information. The information is then transmitted to the G via relay nodes. The VMs communicate valuable information via one or multiple paths, depending on the number of Gs. In other words, each VM has one path that connects one G or shares multiple VMs on one path that connects to one G.

In this study, our aim is to achieve the best system performance according to efficient information extraction to meet real-time constraints. Therefore, we develop a new extraction information algorithm in a suitable reduction ratio that can be achieved in a minimum processing time. Our proposed algorithm is explained in the next section.

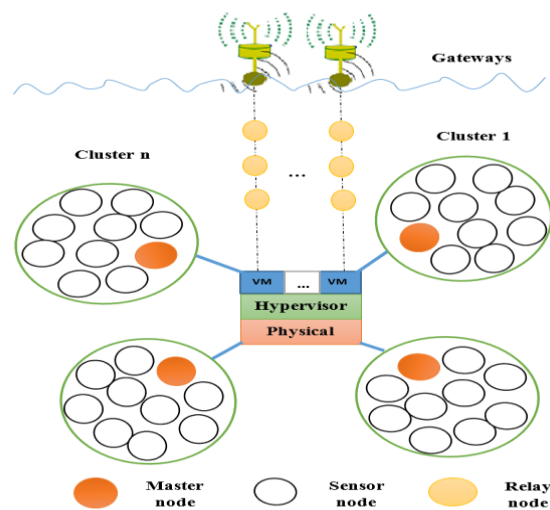


Figure 1: Our proposed architecture.

4 Weighted Records Sample Algorithm

Generally, the challenge of big data applications is the transmission of large volumes of data. Another challenge is the extraction of information due to the limited capability of computer processing (Zhang, 2017). To address these challenges, sampling methods have become a popular technique to extract samples from big data. They have been successfully used in scientific areas such as network traffic, biomedical analysis, and social network analysis.

The sample represents a summary of data sets. For this study, we use the weighted records algorithm to extract the sample and its characteristics. We first extract a random sample of records based on the record's weight. Information is then extracted from the sample, such as the average or the mean of the

features, as illustrated in in Figure 2. We got the original data for seismic monitoring from (Godse, 2017). Then, we reprocess the data, some features represent as text. After that, we generate the same data for different SNs. The steps of this algorithm are as follows:

- Generate random data for different SNs. In our work, the weight of each record depends on the five features: seismic class, hazard, seismic sound, number of pulses, energy.
- Calculate the weight of each feature. The weight of each feature is calculated by dividing the value of each record to the maximum value of feature. Then, resulting value is multiplied by a certain percentage. The percentage of each feature is 20% for each. The total of percentage should be 100%.
- Compute the total weight of each record. The weight of each record is written as:

$$\sum_{i=1}^j \frac{V_i}{M_i} * C_i \quad (1)$$

Where V_i represent as the value of each feature, M_i is maximum value of each feature and C_i the value of the percentage

- Generate the sample of records based on the condition of the record weight value.
- Extract the information from the sample.

Here, the information is represented as sample characteristics such as the average, mean, or variance. The sample characteristics formula can be written as

$$\begin{bmatrix} \text{Information} \\ \text{average} \quad \text{mean} \quad \text{variance} \end{bmatrix} = \begin{bmatrix} \dots \\ \text{Sample} \\ \dots \end{bmatrix} * \begin{bmatrix} \text{single ratio} \\ \alpha_1 \quad \alpha_2 \quad \alpha_3 \end{bmatrix}. \quad (2)$$

The VMs can send the information directly to the Gs or send the suitable sample and its information to the Gs. The goal is to send information and a suitable sample to enhance the error correction at the surface because the UWSNs suffer from data loss during the vertical communication.

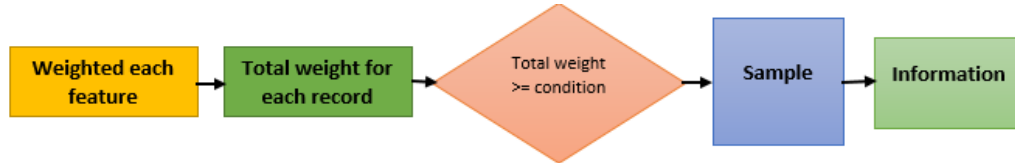


Figure 2: Our proposed algorithm.

5 Results and Discussion

We use the python language to apply our proposed weighted record sample algorithm. We then measure the performance of our architecture via simulation (Albarakati H. A., 2017) (Albarakati H. R., December 2017) and compare it with three approaches. In our approach, we use the weighted record sample algorithm to extract useful information in VMs and then send them via the multipath method. In Magadza approach, the source node (the central node/sink node in the deep layer) divides the data into segments and sends the segments via the multipath method (Magadza, 2015). Zhou approach (Zhuo, 2016), the AUV collects data from the central node/sink node in the deep layer and sends them to the mothership.

In our experiment, we evaluate the performance of our architecture when we apply our proposed weighted record sample algorithm. The performance is evaluated via four factors: bit rates, sample class

size (minimum value of weight record), depth, and lifetime of the network. In our previous work, we proposed data-gathering algorithm which keeps the data load balanced among VMs (Albarakati H. R., 2018). Therefore, we assume the homogenous data in the VMs and each VM has one path. Table 1 shows the size of information and the sample size using the weighed record sample algorithm. The parameters of the experiment are defined in Table 2.

Data gathering is 1MB, 250KB for each VM				
Information and sample class for each VM				
	Information	Sample #1	Sample #2	Sample #3
Condition	Info. of Sample #1	Weighted record $\geq 95\%$	Weighted record $\geq 85\%$	Weighted record $\geq 75\%$
Size	4K	21K	56K	84K

Table 1: Sample class.

To understand the scope of the data extraction class size and bit rate factors, we first set the bit rate to 2500 bps (Khan, 2015), the depth to 2000 m, and the data size to 250 KB for each VM. Figure 3 exhibits the speed-up increase between the two approaches (Our approach and Magadza approach) via different bit rates. Our approach is 22 times faster than the Magadza approach when the VM sends information exclusively, and is 8 times faster than Magadza approach when the VM sends sample#1. Therefore, the speed-up increases when the sample size decreases. The bit rate is also set to 4800 bps according to (Magadza, 2015). We also find that our approach is faster than Magadza approach with different sample sizes.

Figure 4 shows the end-to-end delay for all possible cases to send sample#1 and compares them with the Magadza and AUV approach. In this experiment, we set bit rat to 4800pbs and depth to 2000m. According to the end-to-end delay, our proposed architecture and information extraction algorithm is more efficient than the other approaches (Zhuo, 2016) (Magadza, 2015) .

Parameter	Value		
	Our approach	Magadza approach	Zhou approach
Bit Rate (Magadza, 2015)	4800 bps	4800 bps	-
Number of paths and Gs	4	4	-
Number of VMs	4	0	-
Initial energy of relay node battery E_i	10000 J	10000 J	-
Transmitting and receiving power of the packet [6]	0.1 J	0.1 J	-
Lifetime	-	-	8 hours (Petillo, 2010)
Speed	-	-	2 knot (Zhuo, 2016)
Depth	Unlimited	Unlimited	200-200m (Petillo, 2010)

Table 2: Parameters.

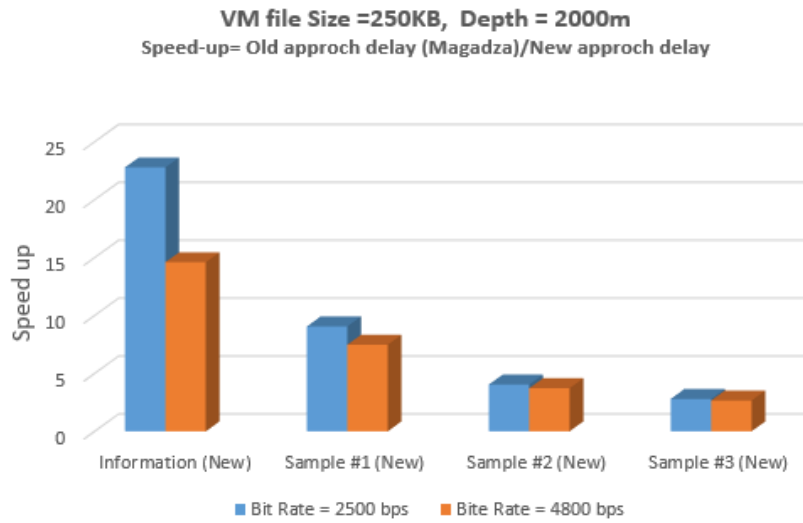


Figure 3: Speed-up between two approaches with different data rates.

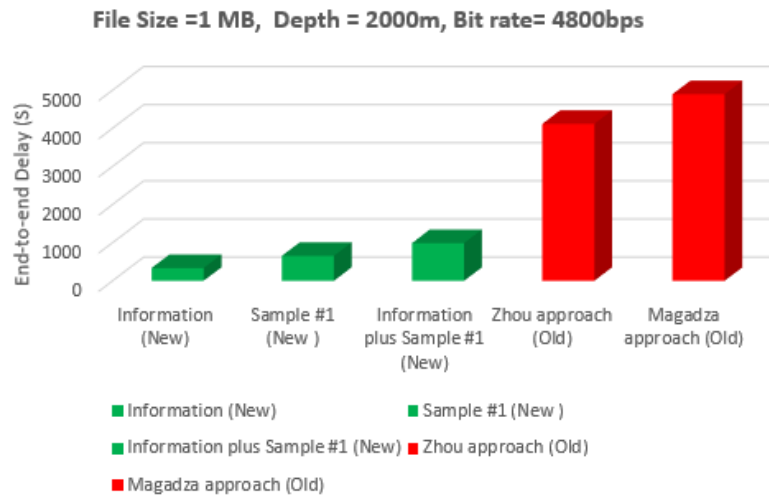


Figure 4: End-to-end delay for all possible cases.

First, we examine the effect of the weighted record sample algorithm on the network’s lifetime. In this experiment, we focus on relay nodes because they send the information/data from the VMs to the Gs. The apparent lifetimes of information, sample#1 and information plus sample#1 approach (our proposition) are longer than that of the Magadza approach and Zhou approach , and the information class (our proposed) significantly increases the network’s lifetime, as illustrated in Figure 5.

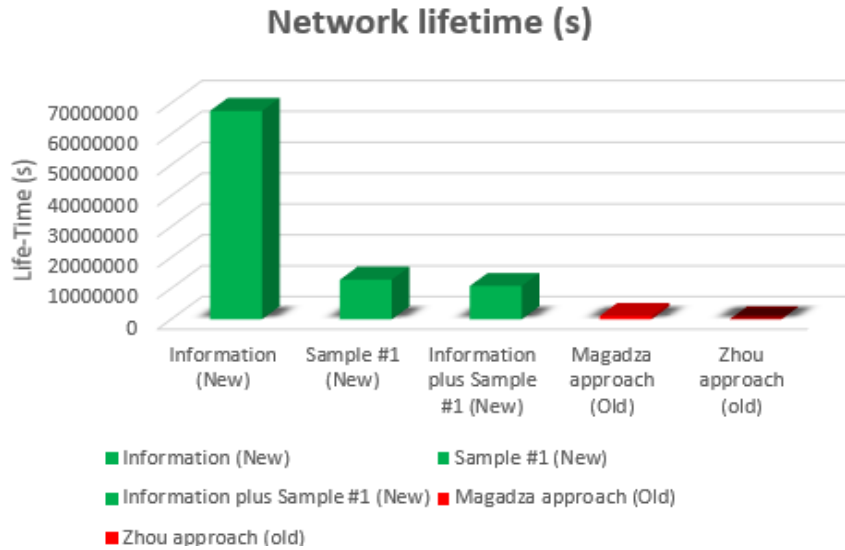


Figure 5: Lifetime of network.

Figure 6 depicts the performance of the two approaches (Our approach and Magadza approach) at depths of 2000 and 10,000 m. In Fig 6 our approach (information) is 13 times faster than Magadza approach. If the depth increases to 10,000 m, the speed-up increases by 32. Therefore, we obtain the highest speed with the longest depth. As a result, the speed-up increases once the depth increases.

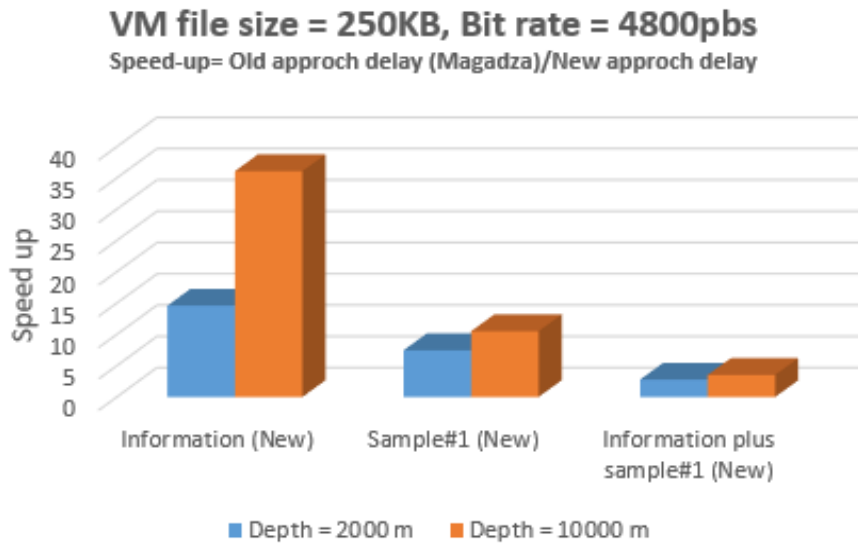


Figure 6: Speed-up between approaches via among depths

6 Conclusions

In this study, we extrapolated our research findings published in (Albarakati H. A., 2017) (Albarakati H. R., December 2017) (Albarakati H. R., 2018). We developed a new algorithm to enhance information extraction in the CC. This algorithm determines the weight of each record and then selects a random sample from the data based on the record's weight. We then extracted the information from the sample. In addition, we evaluated the performance of our system with the new algorithm extraction based on four factors: bit rates, sample size (minimum value of weight record), depths and network lifetime. We presented simulation results to show our algorithm's efficiency.

Our future work will include various algorithms for different applications and comparisons between them.

References

- Albarakati, H. A. (2017). Reconfigurable underwater embedded systems architectures. *In 2017 IEEE Symposium on Computers and Communications (ISCC)*, (pp. 1372-1379).
- Albarakati, H. R. (2018). Real-Time Underwater Computing System. *n 2018 IEEE Symposium on Computers and Communications (ISCC)*, (pp. 00890-00896).
- Albarakati, H. R. (December 2017). Real time underwater embedded system architectures. *In 2017 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, (pp. 156-162).
- Godse, V. K. (2017). Prediction of Seismic Activities in Coal Mines using Machine Learning.
- Heidemann, J. W. (2006). Research challenges and applications for underwater sensor networking. *In Wireless Communications and Networking Conference, WCNC 2006. IEEE, vol. 1*, pp. 228-235.
- Jindal, H. S. (2014). Challenges and issues in underwater acoustics sensor networks: A review. *In Parallel, Distributed and Grid Computing (PDGC), 2014 International Conference, IEEE*, (pp. 251-255).
- Khan, F. A. (2015). Scheduling multiple mobile sinks in underwater sensor networks. *In 2015 IEEE 40th Conference on Local Computer Networks (LCN 2015)*, (pp. 149-156).
- Khan, J. U.-S. (2015). A distributed data-gathering protocol using AUV in underwater sensor networks. *Sensors 15, no. 8 (2015)*, (pp. 19331-19350).
- Khan, J. U.-S. (2016). Data-gathering scheme using AUVs in large-scale underwater sensor networks: A multihop approach. *Sensors 16, no. 10 (2016)*, (p. 1626).
- Magadza, T. B. (2015). VSP: Video streaming protocol for underwater sensor network. *In Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions), 2015 4th International Conference on, IEEE 2015*, (pp. 1-6).
- Petillo, S. A. (2010). Autonomous adaptive environmental assessment and feature tracking via autonomous underwater vehicles. *In OCEANS 2010 IEEE-Sydney*, (pp. 1-9).
- Proakis, J. G. (2001). Shallow water acoustic networks. *IEEE communications magazine 39(11)*, (pp. no. 11: 114-119).
- Zhang, H. L. (2017). Extracting sample data based on poisson distribution. *In Machine Learning and Cybernetics (ICMLC), 2017 International Conference on, vol. 2, IEEE 2017*, (pp. 374-378).
- Zhuo, W. G. (2016). AUV-aided communication method for underwater mobile sensor network. *In OCEANS 2016-Shanghai, IEEE, 2016*, (pp. 1-7).