An Energy Efficient and Traffic Aware Data Fusion Scheme for Water Pollution Monitoring

Emami Yousef*, Javidan Reza and Keshtgari Manijeh

Computer Engineering and IT Department, Shiraz University of Technology, Shiraz, Iran; yousef.emami@ieee.org, javidan@sutech.ac.ir, keshtgari@sutech.ac.ir

Abstract

Background/Objectives: One of the main application of underwater wireless sensor network is water pollution monitoring. Therefore it is proposed to use an energy efficient and traffic aware data fusion scheme (ETDFS) for this kind of application. **Methods:** Data fusion comprises preprocessing and processing steps. In the preprocessing step, a probabilistic data structure entitled bloom filter is used to reduce the number of transmissions and save precious energy. In the processing step, a cost effective data fusion scheme called fuzzy logic is utilized to process the gathered data at sink and determine the level of pollution. **Findings**: ETDFS is superior than other proposed techniques for water pollution monitoring since it reduces the number of transmissions and eliminates redundancy hence the precious energy is saved and water pollution monitoring can be done for longer time. **Application/Improvements:** ETDFS enhance energy efficiency and network lifetime further it reduces network load and loss rate. It can be used in deep waters like sea and ocean.

Keywords: Bloom Filter, Energy Efficiency, Traffic Awareness Data Fusion, Water Pollution Monitoring

1. Introduction

Data fusion is an important technology with rapid growing and has various application in the different realms. Diversity of techniques and architectures make data fusion able to present solutions in different areas¹. The most important applications of data fusion are: Robotic, automatic control of industrial systems, development of intelligent buildings, and medical applications². A fundamental issue in Underwater Wireless Sensor Networks (UWSNs) is the way, the gathered data is processed. When multiple sensors are available, data fusion is a effective tool for enhancing the performance of monitoring system³. In this paper we apply data fusion technology in water pollution monitoring. UWSNs can perform online pollution monitoring (chemical, biological and nuclear) of selected ocean areas in this regard Dissolved oxygen (O₂), Total nitrogen (N), Ammonia (NH₃-N) and Total phosphorus (P) are the most important indicators of pollution. Hence, the goal is to design a low cost, less complicated and effective water pollution monitoring system^{4,5}.

monitoring, undersea explorations, Disaster prevention, assisted navigation, Distributed tactical surveillance, Mine Reconnaissance. Bloom filter utilizes an array of m bits initially all bits set to zero. In order to add an element to bloom filter it hashes k times then the content of corresponding array's indexes generated by the hash operation are set to one, in the rest of this paper the process of adding one element to bloom filter is called bloom. To support membership query for a specific element like y, hash it k times if one of the corresponding bits is zero y is not member of the set. If all of the corresponding bits are one, y is either in the set or is a false positive. Bloom filter may result in false positive, in this case, an element considered to be in the set though it is not. Counting bloom filter is a special version of bloom filter in which each entry in the bloom filter is not a single bit but instead a small counter. Upon inserting an item the relevant counters are incremented and in case of deletion the corresponding counters are decremented⁶.

UWSNs enjoy myriad of application in the diverse arenas the most important ones are Environmental

^{*}Author for correspondence

In this paper an energy Efficient and Traffic aware Data Fusion Scheme (ETDFS) for water pollution monitoring is proposed. The salient feature of the ETDFS is exploiting counting bloom filters in the preprocessing step, this approach reduces the number of transmission and eliminates redundancy hence the precious energy is saved.

The rest of this paper is organized as follows. Existing literature is reviewed in section 2. The preliminary knowledge is introduced in section 3. The proposed method is discussed in section 4. Performance evaluation is presented in Section 5. Finally concluding remarks are presented in section 6.

In this section, the existing literatures on data fusion in terrestrial and underwater wireless sensor networks is presented. Then the novelty of our work is concluded.

Xiong and Xiaohui7 proposed a new prediction-based data fusion scheme using Grey Model (GM) and Optimally Pruned Extreme Learning Machine (OP-ELM) to overcome intrinsic limitations of sensor node, reduce redundant data transmission and save energy. GM-OP-ELM utilizes a dualprediction mechanism to keep the prediction data series at the sink node and sensor node synchronous. Upon completion of sensing interval, sensor make a comparison between the sensed data and the predicted data if the predicted error is subtle and under the threshold, the sensor node does not send the sensed data to the sink node, and the data transmission is canceled to save the energy, which is the object of data fusion. Pinto et al⁸. proposed an approach for implementing data fusion techniques in IEEE 802.15.4. The utilized data fusion technique is Genetic machine learning approach (GMLA) and aims to enhance the communication efficiency by adjusting the sending rate of WSN. GMLA is a feasible solution for real data fusion application. The proposed approach shows performance improvement compared to pure approach based on IEEE 802.15. Jian et al⁹. proposed a new data fusion approach for water quality monitoring using Dempster-Shafer theory. Their experiments indicates the proposed approach can improve evaluation performance in comparison to the approaches utilizing individual sensors. Shaghyegh et al¹⁰. proposed trusted data fusion by using cellular automata (TDFCA) in wireless sensor network, in which cellular automata rules are exploited to execute data fusion, find the cluster head, and the most trusted neighbors for sending the fusion result to the sink. Their result show TDFCA escalates the trust value of the fusion result and enhance network lifetime. Ahmad¹¹ proposed a fuzzy logic-based data fusion technique to support the maritime surveillance process. A real data set is utilized to evaluate the effectiveness of the proposed data fusion algorithm. Their results demonstrate computation complexity is not high and is in acceptable range. Larios et al¹². proposed an instance of data fusion based on Self-organizing Map (SOM) applied to a WSN, intending to increase lifetime. Tang et al¹³. proposed an information fusion framework to recognize underwater target type and infer diverse target characteristics from sensory information using dynamic Bayesian network.

Bloom filters are frequently utilized in the realm of network security. Wireless networks utilize bloom filters for authentication, anonymity and privacy-preserving firewalling, tracebacking, misbehavior detection, replay attack detection and node replication detection. Wired networks use bloom filters for string matching, IP tracebacking, spam filtering and email protection, DoS and DDoS attacks detection and anomaly detection¹⁴. Denh et al¹⁵. proposed Coordinated Packet Traceback protocol (CAPTRA) for wireless sensor networks. In CAPTRA each sensor dedicate a subtle memory for the bloom filter. When a packet passing a network, each forwarding sensor, records the packet in its bloom filter later the bloom filter will be used to rebuild the attack graph.

Compared with the above-mentioned related works, the contributions of our paper can be listed as follows:

- A novel data fusion approach (ETDFS) based on bloom filter is proposed for water pollution monitoring.
- A novel, light and Indispensable part of ETDFS is its preprocessing step in which bloom filter reduces the number of transmission and redundancy without posing any extra overhead unlike some of the discussed works. Bloom filter is fully compatible with sparse deployment of UWSNs.
- Simulation experiments are conducted to evaluate the performance of ETDFS. The results demonstrate that ETDFS can improves lifetime and energy efficiency and reduces network load and loss rate.

2. Literature Review

In this section, the existing literatures on data fusion in terrestrial and underwater wireless sensor networks is presented. Then the novelty of our work is concluded

3. Premilinaries

3.1 Underwater Propagation Model

In this paper, the Urick path-loss formula⁴ is utilized to characterize the acoustic channel:

$$TL (d, f) = \chi \log (d) + \alpha (f) d + A$$
(1)

Where d is internode distance and f is operating frequency. The term χ stands for the *geometric spreading* and varies depending on the depth of the water, it is spherical spreading in our case. Last term, A is the transmission anomaly.

According to⁴, the propagation speed of acoustic signals is determined by:

$$q(z, S, t) = 1449.05 + 45.7 \cdot t - 5.21 \cdot t^{2} + 0.23 \cdot t^{3} + (1.333 - 0.126 \cdot t + 0.009 \cdot t^{2}) \cdot (S - 35) + 16.3 \cdot z + 0.18 \cdot z^{2}$$
(2)

Where t = T/10 (T is the temperature), S is salinity in ppt, and z is the depth in km.

The above equation is an effective tool to determine the propagation speed.

3.2 Underwater Channel Model

The reason behind severe degradation of the acoustic communication signal is multi-path propagation since it generates Intersymbol Interference (ISI). But, the deepwater acoustic channel is not affected by multiple path. Instead, a higher attenuation is present due to the spherical spreading of the acoustic signal. As shown in (2), depth and temperature are key factors in determining the propagation speed of acoustic signals. Hence, the propagation path depends on the position of the transmitter and also thermal structure of the particular medium⁴.

4. The Proposed Method

ETDFS comprises preprocessing and processing steps. The goal of preprocessing is to diminish the number of transmissions using counting bloom filter so that energy consumption to be reduced. Processing step is done in the sink and exploit a data fusion technique entitled fuzzy logic for water pollution monitoring. Figure 1. (All figures can be found at the end of the paper) Depicts the general structure of the proposed method. In the rest of this section these steps are deeply investigated.



Figure 1. Flow chart of the proposed algorithm.

4.1 Preprocessing Step

In this step, 32 bit bloom filters are used and the number of considered hash functions is set to 3. Each front-end node senses the underwater environment for the target features, in our case NH₂-N, P, N and O₂. Moreover, adds the sensed data for each feature to a dedicated bloom. Then, sends the built bloom filters toward the middle nodes. The middle nodes also sense and bloom the target features. Meanwhile, the middle nodes consolidate the received bloom filters. In the end the middle nodes integrate the consolidated bloom filters with its own bloom filter and send the result toward the sink using DBR routing protocol. The sink unbloom the received bloom filters and convert them to the real data so that the fuzzy algorithm to be executed on them and level of pollution to be determined. The proposed approach, convert multiple packet to one packet using bloom filter. Hence, the number of transmissions and redundancy is reduced. But, the conventional method transmit each packet without blooming. Figure 2. Illustrates this step in detail.

4.2 Processing Step

The utilized data fusion technique in the processing step of the proposed algorithm is fuzzy logic. The fuzzy technique does not require complex mathematical equation and system modeling and its salient features are lower costs and better performance¹⁶. In the fuzzy logic, the



Figure 2. Flow chart of the preprocessing step.



Figure 3. Fuzzy structure with four inputs and one output.

crisp values are fuzzified then the fuzzified values are processed by inference engine which consists of rule base (a series of IF-THEN rules) and diverse methods to inference the rules (Mamdani in our case) finally the fuzzy values are defuzzified and crisp output is generated. Based on the achieved data in the preprocessing step, the fuzzy logic determines the quality of the water that depends



Figure 4. Membership graph for the inputs and the output.

Table 1. If-Then Rules.

no	Antecedent				Consequent
	0,	Р	N	NH3-N	Result
1	L	L	L	L	SEMI P
2	L	L	L	Н	SEMIP
3	L	L	Н	L	SEMIP
4	L	L	Н	Н	Р
5	L	Н	L	L	SEMIP
6	L	Н	L	Н	Р
7	L	Η	Н	L	Р
8	L	Н	Н	Н	Р
9	Н	L	L	L	С
10	Н	L	L	Н	С

on the dissolved oxygen O_2 , total nitrogen N, ammonia NH3-N and total phosphorus P. Figure 3. illustrates the fuzzy approach with four input parameters O_2 , N, NH3-N and P, and an output, with universal of discourse [0...19], [0...100], [0...45], [0...30] and [0...100], respectively. Our method uses five membership function for each input and output parameter as shown in Figure 4. Table 1. shows some of the IF-THEN rules used in the fuzzy approach. As an example, IF O_2 is high and P is low and N is low and NH3-N is high THEN output is clear. In the end, the defuzzification finds a crisp output value. Defuzzification is done using centre-of-maxia method¹⁷ given by

Output =
$$\frac{x_1\mu_1 + x_2\mu_2 + \dots + x_n\mu_n}{\mu_1 + \mu_2 + \dots + \mu_n}$$
(3)

Where \mathbf{x}_n is the numerical value and is the degree of membership.

5. Performance Evaluation

To evaluate the efficiency of ETDFS in terms of energy efficiency, lifetime, network load and loss rate simulation experiments are conducted to compare the performance of ETDFS based on bloom filter with the conventional method. In conventional method blooming is ignored, each front-end node sense data in the determined times and send them toward the middle nodes, then the middle nodes receive and buffer these data, in special times the buffered data is extracted and sent for the sink using DBR. Simulation results have shown that ETDFS reveals better performances than conventional method.

The following metrics are used for the performance evaluation:

Energy efficiency: Total energy consumption by sensor is not a good criteria. A scenario may deliver more data consequently more energy will be consumed so for a better comparison, energy efficiency is used as criteria and defined by

Life time: The *time* until at least 10 percent of the nodes are drained of their *energy*.

Network load: refers to the amount of data (traffic) being carried by the network.

Loss Rate: The amount of lost information by the network (transmitted data not delivered to destination due

to collision). In other word loss rate is equal to1-packet delivery ratio.

5.1 Simulation Setup

The simulations are carried out in Aqua-Sim. Aqua-Sim can simulate acoustic signal attenuation and packet collisions in underwater sensor networks. In addition, Aqua-Sim can easily be integrated with the existing codes in NS-218. The 200 nodes are randomly scattered in a 1000*1000*1000 area in the depth of 200 m. The coordinates of sink is (200, 200, 0) and there is only one sink. All sensor nodes have the same initial energy 20J. In the data processing interval, the sink processes data each 50 second using fuzzy algorithm. The energy consumption parameters are set according to the UWM1000 LinkQuest Underwater Acoustic Modem¹⁹. The exploited MAC and routing protocol are Broadcast MAC and DBR respectively. Type of antenna and transmission range are akin to²⁰, each sensor is equipped by an omnidirectional transceiver and the transmission range of each sensor is 100 meter. Direct communication is established between two nodes if they are within their communication range. Detailed simulation settings are list in Table 2.

In data processing interval, data are processed using fuzzy logic, in data gathering interval the middle nodes bloom the received packets and send them toward sink,

Parameter	Value		
Network scale	1000m*1000m*1000m		
Location of sink	(200,200,0)		
Number of nodes	200		
Simulation time	2000 s		
Acoustic link	17.8 kbps		
Sensor type	LinkQuest UWM1000		
Initialized energy	20 J		
txPower	2 w		
rxPower	0.75 w		
Propagation speed	1500 m/s		
Operating frequency	35 kHz		
MAC protocol	Broadcast MAC		
Routing protocol	DBR		
Sensing interval	3 to 15 s		
Data processing interval	50 s		
Number of front-end nodes	20		

Table 2.Simulation Settings

finally in sensing interval, the front-end nodes and also the middle nodes sense the chemical features of water.

5.2 Simulation Results

All obtained results in the figures have been repeated 10 times and their average have been presented. The confidence interval can reach 95%. To obtain the following results, sensing and sending interval vary from 3 to 15 second, gathering interval in the middle nodes is 20 second and fuzzy processing interval is 50 second. As it can



Figure 5. Energy Efficiency.



Figure 6. Lifetime.



Figure 7. Network Load.



Figure 8. Loss Rate.

be seen in Figure 5, ETDFS show an improvement from energy efficiency point of view than the conventional method, since the ETDFS combine multiple packet in one using bloom filters in the middle nodes thus energy consumption is reduced and energy efficiency is improved. Figure 6, depicts the ETDFS has longer lifetime than the conventional method since the number of transmitted packets in the middle nodes is reduced therefore energy consumption by each sensor is diminished, less energy consumption in sensor result in increased lifetime. Figure 7, demonstrates the amount of generated traffic in the ETDFS is clearly less than the conventional method, therefore traffic load in the ETDFS is also less. The Figure 8, indicates the loss rate in the ETDFS is less than its primary counterpart obviously the decreased number of the transferred packets reduce the number of collisions.

6. Conclusion

One of the main characteristics of UWSNs is that the network lifetime is highly related to the number of transmission. To efficiently reduce the number of transmission and to prolong the overall lifetime of UWSNs, weproposed ETDFS, a new energy-efficient and traffic aware data fusion scheme for water pollution monitoring by using a combination of both FUZZY approach and Counting bloom filter. In the preprocessing step, counting bloom filters are exploited to diminish the number of transmission so that energy consumption to be reduced.

7. Acknowledgment

Yousef Emami would like to express his sincere gratitude to Dr. Reza Javidan, his co-author, for persevering with him as supervisor through out the time it took him to complete this research and write the dissertation.

8. References

- Thomas C. Sensor Fusion-Foundation and Applications. 2011.
- 2. Liggins ME, et al. CRC Press: Handbook of Multisensor Data Fusion. 2009.
- Nakamura EF, et al. Information fusion for wireless sensor networks: Methods, models, and classifications. ACM Comput. Surv. 2007; 39(3):1–9.
- 4. Ian MCV, Akyildiz F. Wireless Sensor Networks. 2010.
- 5. Poehls DJ. Encyclopedic Dictionary of Hydrogeology. 2009.
- Broder A, et al. Network Applications of Bloom Filters: A Survey. Internet Mathematics. 2002; p. 636–46.
- Luo X, Chang X. A novel data fusion scheme using grey model and extreme learning machine in wireless sensor networks. International Journal of Control, Automation and Systems. 2015; 13(3):539–46.
- Pinto AR, et al. An approach to implement data fusion techniques in wireless sensor networks using genetic machine learning algorithms. Information Fusion. 2014; 15(Special Issue):90–101.
- Zhou J, et al. Multisensor Data Fusion for Water Quality Evaluation Using Dempster-Shafer Evidence Theory. International Journal of Distributed Sensor Networks. 2013; Article ID 147419:1–6.
- 10. Jaberi S, et al. Trusted data fusion by using cellular automata in wireless sensor networks. Perth, Australia: pre-

sented at the Proceedings of the 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference. 2011; p. 145–51.

- Gad AS. New Cairo: A fuzzy logic-based multisensor data fusion for Maritime Surveillance in Radio Science Conference.2009; p. 1–10.
- Larios JBDDF, Rodriguez G, Sevillano JL, Molina FJ, Leon C. Energy efficient wireless sensor network communications based on computational intelligent data fusion for environmental monitoring. IET COMMUNICATION. 2012; 6(14):2189–97.
- Tang Z, et al. Multi-sensor data fusion for underwater target recognition under uncertainty, Hangzhou, China: 2010 2nd International Conference on Information Science and Engineering (ICISE). 2010; p. 1315–18.
- Geravand S, Ahmadi M. Bloom filter applications in network security: A state-of-the-art survey. Computer Networks. 2013; 57(18):4047–64.
- Denh S, Lichun B. CAPTRA: coordinated packet traceback. Nashville: IPSN 2006, The Fifth International Conference on Information Processing in Sensor Networks. 2006; p. 152–59.
- Gad AS. A fuzzy logic-based multisensor data fusion for Maritime Surveillance — real data testing. National, New Cairo: NRSC 2009, Radio Science Conference. 2009; p. 1–10.
- 17. Runkler TA. Selection of appropriate defuzzification methods using application specific properties. Fuzzy Systems, IEEE Transactions on. 1997; 5(1):72–79.