This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2011 proceedings

Minimum-Latency Aggregation Scheduling in Underwater Wireless Sensor Networks

Zuodong Wu   Chen Tian   Hongbo Jiang   Wenyu Liu
Dept. of EIE, Huazhong University of Science and Technology, China

***Abstract*—Underwater Wireless Sensor Networks(UWSNs) can enable a broad range of applications; data aggregation is a fundamental task in such multi-hop wireless sensor networks. To the best of our knowledge, none of existing research works have addressed the interference-free data aggregation scheduling problem in UWSNs. In this paper, we formally define the data aggregation model in UWSNs. We propose a realistic aggregation scheduling scheme together with its theoretical latency bound $R_h(C(\Delta-1)+D)$, where R_h and Δ are the hop radius and the max degree of the network respectively while C and D is a constant. Specifically, we introduce the concept of *Virtual Slot* to efficiently exploit multiplexing opportunities of time domain. Compared with naively adapted terrestrial algorithms, the evaluation results show that our proposed algorithm achieve far better performance especially when the packet size is small or the node density is high.**

I. Introduction

Underwater Wireless Sensor Networks(UWSNs) can enable a broad range of applications. Compared with ground-based sensor networks, one fundamentally different issue in underwater circumstance is the communication delay: radio and optical waves are not suitable for underwater communications, and acoustic communication is the typical physical layer technology in underwater networks [4]; propagation delay is one of the central factors of Underwater Wireless Sensor Networks.

In multi-hop wireless sensor networks, a primitive task is to gather data from all sensors to a distinguished $sink$ node. Quite often every intermediate node would merge received data with its own record according to some aggregation functions into a single packet of fixed-size. *Minimum-Latency Aggregation Scheduling*(MLAS) addresses the problem of minimizing the data aggregation latency of the whole convergecast process [8].

Due to the existence of mutual interference in terrestrial sensor networks, synchronized aggregation scheduling algorithms are designed where all communications proceed in synchronous time slots [5], [6], [7], [8]. Such a data aggregation scheduling includes both a spanning inward tree rooted at the $sink$ and a link scheduling of this spanning routing tree *ruled by*:

1) each node can transmit at most one packet of a fixed size in its specified time slot;
2) a node can transmit only after all its assigned children have completed their transmissions to itself;
3) all transmissions assigned in the same time slot are interference-free.

Due to the long propagation delay of acoustic signals, current terrestrial aggregation scheduling schemes are not suitable for UWSNs. We provide a simple example, as shown in Figure. 1, to demonstrate the effect of the high propagation delay in UWSNs over aggregation scheduling schemes. Figure. 1(a) represents the topology of a network: all nodes a, b and c would send their data to be aggregated at node v; the distances from them to node s is 45m, 30m and 15m respectively. Suppose the length of both a time slot and a data frame is 0.01s; we also assume the underwater acoustic speed is fixed as 1500m/s, hence data from a, b and c would experience 3, 2 and 1 slot propagation delay respectively. In the terrestrial scenario, the interference-free schedule 1 is shown in Figure. 1(b) that node a sends at slot 1; node b sends at slot 2; and node c sends at slot 3. If all nodes send at the same time slot, the frames would collide with each other in the receiver side (scheduling 2 in Figure. 1(c)). However, the schedule for the same topology in UWSN is different due to the propagation delay. Schedule 1 causes collisions for underwater scenario, as shown in Figure. 1(d), while schedule 2 is able to aggregate data from node a, b, and c without conflicts, as shown in Figure. 1(e), because the reception time of the three packets is naturally separated.

To the best of our knowledge, none of existing research works on underwater networks have addressed the interference-free data aggregation scheduling problem. Terrestrial scheduling involves assigning time slots to transmission links; the vertex coloring theory is widely used. In underwater scenarios, assigning a transmission time to source node would also fix an occupied period, which equals the transmission time length of the frame, at the receiver node. The contributions of our work are as follows:

- We formally define the data aggregation models in UWSNs, together with the formulation of Minimum-Latency Aggregation Scheduling(MLAS) problem.
- We develop a virtual slot based approximation algorithm which efficiently exploits multiplexing opportunities of time domain and present its theoretical bound.
- We show in the evaluation section that our proposed algorithm outperforms the algorithm naively adapted from a state-of-art algorithm for the terrestrial scenario, especially when the packet size is small or the node density is high.

The rest of this paper is organized as follows. We formally define the aggregation scheduling model in Section II, together

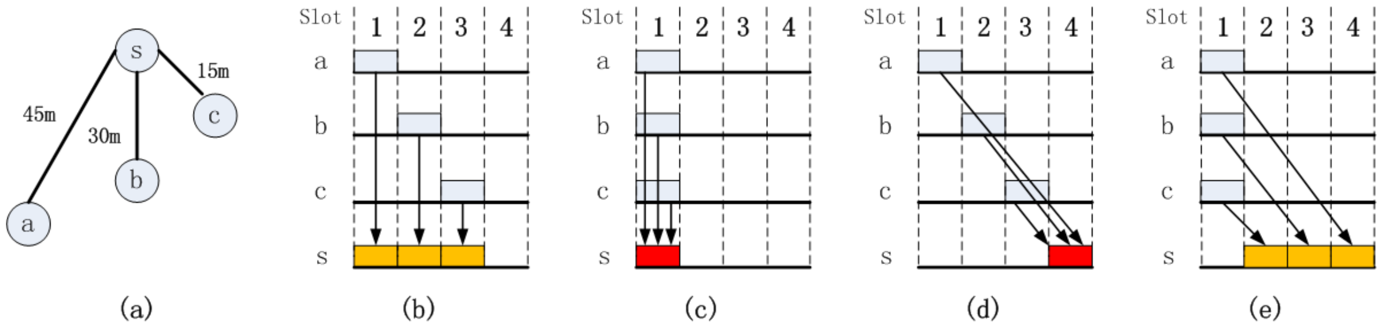


Fig. 1. (a)Network topology (b)(c) schedule 1 and 2 in terrestrial scenario (d)(e) the same schedules in underwater scenario.

with some preliminary knowledge. In section III we review related works. The approximation algorithm based on Shortest Hop Tree(SHT) is described in Section IV. We evaluate the proposed scheduling algorithm by simulations in Section V. Finally, we represent our conclusion in Section VI.

II. Background

A. Problem Formulation

1) System Model: In this paper, we study the MLAS problem under the following system model for UWSNs. All nodes are located in an Euclidean plane and are each equipped with an omnidirectional acoustic antenna. Each node has a fixed transmission/interference radius denoted by D. The communication/interference range of a node u is a disk centered at u of radius D. We assume that the underwater acoustic speed $Speed$ is relatively stable as the measured propagation delay variation between two sensor nodes in UWSNs is at most one percent [3]. Such interference model is referred to as the *protocol interference model* [9] and is widely used because of its generality and tractability.

2) Data Aggregation Model: One of the main tasks of sensor nodes is to collect data and transmit them back to sink z by convergecast. Data aggregation problem in UWSNs must be *NP-hard* inferred from [5]. An approximation of MLAS is a sequence of sets of senders $\{S_1, S_2, \ldots, S_l\}$ satisfying following conditions:

- $S_i \cap S_j = \emptyset, \forall i \neq j$;
- $\cup_{i=1}^{l} S_i = V - \{z\}$;
- Each node of $\cup_{i=1}^{k} S_i$ transmits at the beginning of the kth slot, for all k = 1, 2, , l;
- Unlike in terrestrial scenarios, slots are allowed to overlap with each other and own different sizes;
- Data are aggregated from $\cup_{i=1}^{k} S_i$ to $\cup_{i=1}^{k+1} S_i$ for all k = 1, 2, , l-1, and at last data are aggregated from $\cup_{i=1}^{l} S_i$ to z in the lth slot.

Its object is to minimize the aggregation time.

B. Preliminaries

The graph distance between any two nodes u and v in G is denoted by $dist_G(u, v)$. The radius of G with respect to a specific node $v \in V$ is denoted by $Rad(G, v)$. Now, fix a node $s \in V$. The depth of a node v (with respect to s) is $dist_G(s, v)$. For each $0 \leq i \leq Rad(G, s)$, the set of nodes in V of depth i is referred to as the i-th layer of G.

The real distance between any two nodes u and v in G is denoted by $D(u, v)$; the propagation delay between u and v denoted by $P(u, v)$ is $D(u, v)/Speed$. Set $P_u = \max P(u, v)$ and $P_l = \min P(u, v)$, for all $u, v \in V$. The transmission delay of the aggregation packet is denoted by F.

III. Related Work

Since the 1980s, algorithms for WSNs have been studied extensively. Data aggregation, however, is relatively new, especially when the facts are considered that practical distributed algorithms is far from mature, and that scheme for UWSNs has not yet been investigated.

There are now two main types of state-of-art centralized algorithms for minimizing data aggregation latency in terrestrial WSNs, both of which are tree-based algorithms and perform in two independent phases.

A. Greedy Scheduling Scheme

The most representative existing algorithm is the Shortest Data Aggregation(SDA) algorithm presented by Xujin Chen *et al* in [5] which guarantees a $R_h(\Delta - 1)$ upper bound. SDA constructs a Shortest-Path-Tree (SPT) T_0 in the first phase with the sink node as its root; after that, the scheduling is iteratively implemented. In rth round($r = 1, 2, 3...$), SDA picks *senders* from all leaf nodes in T_{r-1}, guided by the interference-free principal; then it eliminates the senders from T_{r-1} to obtain T_r. It ends when T_r only contains the root node.

B. Layer by Layer Scheduling Scheme

Unlike the greedy scheduling algorithm, a layer by layer scheduling first divides all nodes into layers based on Breadth First Search(BFS). It schedules senders layer by layer, subject to the interference-free principal as well; to be more specific, it would not choose nodes as senders unless all nodes in lower layers have already been scheduled. Main latest works of the scheme are like the First-Fit algorithm proposed by Scott C.-H. Huang *et al* in [6] and the Sequential Aggregation Scheduling(SAS) proposed by Pengjun Wan *et al* in [8]. Both of the two allocate nodes to layers by finding a Connected Dominating Set(CDS) of the constructed BFS tree and assigning all dominatees(i.e. nodes not in CDS) to the deepest layer.

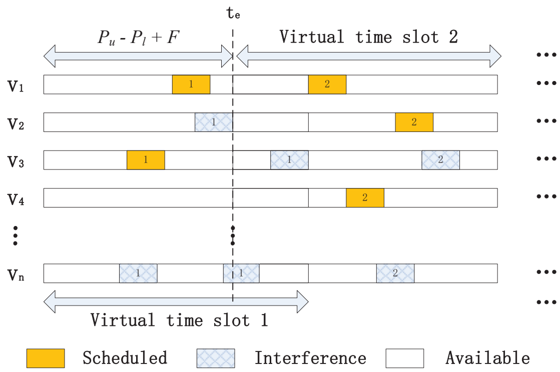


Fig. 2. An example of Spatial-Temporal-Confliction-Table

Unfortunately, the interference-free judgement implementation in the First-Fit algorithm is problematic, which was also found by other researchers [7]. As a successor, SAS achieves a linear theoretical bound $15R_h + \Delta - 4$.

IV. Shortest Hop Tree Aggregation

In this section, we present a layer by layer scheduling algorithm in which a Shortest Hop Tree(SHT) is utilized in the tree construction phase. SHT also subjects to the scheduling rules of the spanning routing tree presented in Section I. In the scheduling phase, we propose two schemes, a basic one and an improved one.

A. Tree Construction

For each $0 \leq i \leq Rad(G, z)$, we denote nodes in the i-th layer of G as $V_i(V_0 = z)$. For each $1 \leq i \leq R_h$, each node $v \in V_i$ sets its parent $p(v)$ to be the node with the smallest ID in V_{i-1} which is adjacent to v. For each $1 \leq i \leq R_h$, A_i denotes the set of links from the nodes in V_i to their parents.

Our aggregation schedule proceeds in R_h rounds, with the $(R_h + 1 - i)$-th round devoted to the links in A_i for each $1 \leq i \leq R_h$. Specifically, we sort all links in A_i in the increasing order of heads (i.e., receiving nodes) and break the ties with the increasing ordering of tails (i.e., transmitting nodes). Such ordering is referred to as *ID-lexicographic* ordering of A_i.

B. Basic Scheduling Algorithm

We introduce the concept of *Virtual Slot*: as each transmission can last no more than $P_u + F$ time, we deliberately set this as bound of the unit of scheduling. On one respect, all links scheduled to transmit at the beginning of a virtual time slot would not interference with transmissions in forthcoming slots; hence the problem is similar to that in terrestrial scenarios: minimizing the used number of virtual time slots. On the other respect, the conflict is not judged by the conflict graph but by the Spatial-Temporal-Confliction-Table, and adjacent slots are allowed to overlap with each other in order to efficiently exploit the multiplexing opportunities of time domain. Due to the definition of data aggregation, in each round the source set U and destination set V are disjoint. An element of Spatial-Temporal-Confliction-Table is shown in Figure. 2:destination nodes $v \in V$ are listed according to their IDs, each has a complete $P_u + F$ time to be scheduled.

The first-fit coloring of A_i is shown as in Algorithm. 1. Each link in A_i with color j is scheduled in the jth time-slot of the $(R_h + 1 - i)$th round.

Algorithm 1 First-Fit Coloring

```
INPUT: A_i
OUTPUT: A scheduling of A_i, S_1, S_2, ...
X ← Ø, j ← 1;
while X ≠ A_i do
    S_j ← Ø;
    while there exists an unscheduled link A_ik in A_i whose
    reception would not be interfered by any link in S_j do
        Add A_ik to S_j;
    end while
    Color all links in S_j with color j;
    Add S_j to X;
    j = j + 1;
end while
```

C. Analysis

Theorem 1: A scheduling of SHT produces an interference-free scheduling in each round.

Proof: Suppose $(u, p(u))$ and $(v, p(v))$ belong to the *potential interference edges set* of each other, $(u, p(u))$ precedes $(v, p(v))$ and $(u, p(u))$ is already scheduled, we are checking $(v, p(v))$ now. Apparently $p(v) \neq p(u)$ as $p(u)$ is already scheduled to receive. If $(v, p(v))$ is found to be conflicted with $(u, p(u))$ in $ST - CT(A_i)$, since $p(u)$ has smaller ID than $p(v)$, by the choice of parent, v is not adjacent to $p(u)$. Therefore, the scheduled links would be interfered by link $v - p(v)$. Thus, if $(v, p(v))$ is found to have unoccupied temporary space, its scheduling would not bring any new interference. ■

Lemma 1: The latency of a round of algorithm SHT is at most $N * (F + P_u - P_l) + P_l$, where N is the number of slots of the round.

Proof: Set transmission of the ith slot of the kth round starts at t_i^k and denote P_{iu}^k and P_{il}^k as the shortest and the longest propagation delay of the slot respectively. Recall that F is a constant which represents the packet transmission delay. As shown in Figure. 2, the completion time of the longest link of the ith slot is $t_i^k + P_{iu}^k + F$, while the arriving time of the shortest link of the $(i+1)$th slot is $t_{i+1}^k + P_{(i+1)l}^k$. A sufficient condition ensuring conflicts between the ith and the $(i+1)$th slot not to happen is:

$$t_{i+1}^k + P_{(i+1)l}^k >= t_i^k + P_{iu}^k + F$$

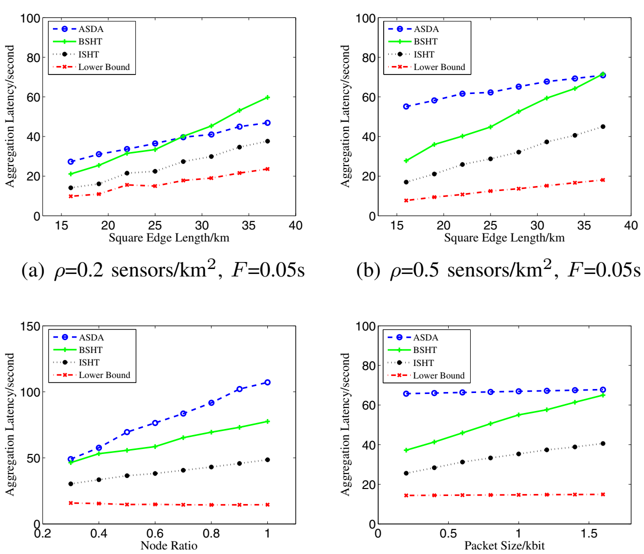
That is $t_{i+1}^k - t_i^k >= P_{iu}^k - P_{(i+1)l}^k + F$. Minimum latency is achieved as long as the equality is guaranteed. Hence, the minimum total latency L of a round with N slots satisfies

$$
\begin{aligned}
L &= t_N^k + P_{Nu}^k + F - t_1^k \\
&= \sum_{i=1}^{N-1} (t_{i+1}^k - t_i^k) + P_{Nu}^k + F \\
&= \sum_{i=1}^{N} P_{iu}^k - \sum_{i=2}^{N} P_{il}^k + N * F \\
&\le N * P_u^k - (N-1) * P_l^k + N * F \\
&\le N * P_u - (N-1) * P_l + N * F \\
&= N(F + P_u - P_l) + P_l
\end{aligned} \tag{1}
$$

In Formula.1, P_u^k and P_l^k are the max and min propagation delay in the kth round respectively. The equality holds, when the size of each slot is fixed as $P_u + F$. It should be noted the prerequisite that SHT is a layer by layer algorithm so that senders of a slot would not be receivers of another slot if the two slots belong to the same round in an aggregation process. In addition, at $L + t_0^k$, all packets sent by nodes in the $R_h - k + 1$th layer have been completely received and would not affect the data transmission of lower layers. These guarantee that the deduction would not contradict rule 2) of the spanning tree method presented in Section I. ■

Theorem 2: Algorithm SHT produces an aggregation schedule with latency at most $R_h(C(\Delta - 1) + P_l)$, where $C = P_u - P_l + F$ is a constant.

Proof: Based on the Lemma 1 and the fact that there are R_h rounds in the aggregation process, all need to be proved is that for any round, the number of slots $N \le (\Delta - 1)$.



(a) ρ=0.2 sensors/km^2, F=0.05s

(b) ρ=0.5 sensors/km^2, F=0.05s

(c) L=30km, F=0.05s

(d) ρ=0.5 sensors/km^2,L=30km

Fig. 3. Four Sets of Evaluations

Except $V_0 = z$, each node in V_{i-1} has at most $\Delta - 1$ children in V_i. For each head, either at least one of its children is scheduled; or its scheduling would cause a conflict which means at least one of its neighbors has been scheduled already, hence the degree would reduce at least 1 for each iterations. After at most $\Delta - 1$ iterations, all children in V_i are scheduled. Hence each round has at most $\Delta - 1$ time slots, *e.i.* $N \le (\Delta - 1)$. ■

D. Improved Scheduling Algorithm

According to the deduction of Lemma 1, if we allow the virtual slot size varies and utilize different P_{mu}^k and P_{ml}^k for different slots, we can achieve higher efficiency in utilizing our multiplexing time scheme and thus shorter aggregation latency than the basic algorithm. Set N_k be the number of slots in the kth round, the latency of this *Improved SHT*(ISHT) L_{Impr} will be

$$L_{Impr} = \sum_{k=1}^{R_h} \sum_{m=1}^{N_k} P_{mu}^k - \sum_{k=1}^{R_h} \sum_{m=2}^{N_k} P_{ml}^k + F * \sum_{k=1}^{R_h} N_k \tag{2}$$

According to the Formula 1 and 2, it can be directly inferred that the data aggregation latency gets lower when

1) the packet size gets smaller;
2) the difference between P_{mu}^k and P_{ml}^k gets smaller for any slot m in any round k.

V. Evaluation

A. Setup

According to [10], the acoustic band under water is limited due to absorption, and most acoustic systems operate below 30kHz.

We implemented adapted SDA(ASDA) to compare its performance with SHT. We randomly deploy N sensors into a square region of edge length L; the density of nodes is determined by $\rho = N/L^2$ and sensors have the same transmission range λ. Our simulation is implemented assuming that *4, 8PSK* is used. According to [10], the data rate R_d and λ are 20kbps and 3.5km respectively. We have also calculated the transmission latency from the remotest sensor to the sink as a referenced lower bound.

The sink is located in the center. All comparisons are fairly conducted on the same graph, and on each graph, the data are aggregated from the same set of nodes to the same sink. For each set of parameter configurations, we perform comparisons with 10 random graphs and present the averaged result.

B. Fixed Node Density

In the first set of experiments, we kept node density fixed as $\rho = 0.2$ sensors/km^2 and the packet size fixed as 1kbit, thus *the transmission delay of a packet* F is 1kb/20kbps = 0.05s. Then we varied L from 16km to 37km, 3km apart. In the second set of experiments, all conditions are the same as the previous one except for that the value of ρ is changed to 0.5 sensors/km^2.

The results of these two sets of experiments are reflected in in Figure. 3(a) and Figure. 3(b) respectively. Considering the two sets separately, we can see that ISHT achieves better performances than BSHT and ASDA for all values of L and that the aggregation latency obtained by ISHT is close to the lower bound. When put them together, it is obvious to find that ISHT outperforms ASDA much more when ρ changed from 0.2 to 0.5 sensors/km^2, intuitively, it is because

1) greater node density brings more leaves to be considered for a round of ASDA, however the number of sensors scheduled to send in the round would not vary much, hence more rounds are needed when ρ is increased;
2) the number of rounds or layers will not vary great for the layer by layer scheme based algorithm SHT, due to the invariance of λ. Moreover, the time space in a round of ISHT is probably utilized so efficient that a proper increment of sensors' number to be scheduled in the round would result in few latency rises.

This variance will be reflected more clearly in the following subsection.

C. Variable Node Density

We kept L and F fixed as 30km and 0.05s respectively and varied the value of ρ from 0.1 to 0.8, 0.1 apart. According to Figure. 3(c), except when ρ is small, the performance of ISHT is much better than ASDA; even BSHT outperforms ASDA. These results accord with that in Section V-B.

D. Variable Packet Size

In this last set of experiments, we kept ρ fixed as 0.5 sensors/km^2 and L fixed as 30km. In Figure. 3(d), we presented the aggregation latency needed when packet size varied from 0.2kbit to 1.6kbit, 0.2kbit apart. The result accords with the first inference in Section IV-D that the aggregation latency gets lower as the packet size gets smaller due to the reduction in the probability of conflicts between any two packets. The reason why the aggregation latency achieved by ASDA is nearly constant lies in that the packet size changes would have no influence on the network topology.

E. Interval Estimation

In order to measure the degree of reliability of our experiments, we implemented an interval estimation of aggregation latency on the condition ρ=0.2 sensors/km^2, F=0.05s, λ=3.5km and L=25km. The result is shown as in Figure 4. The mean value of the results from 1000 randomly generated WSNs is μ=25.02km, while the standard derivation is σ=5.12km; 79.4% of the results lie in the interval $\mu \pm \sigma$, 94.8% lie in $\mu \pm 2\sigma$, while 97.9% lie in $\mu \pm 3\sigma$.

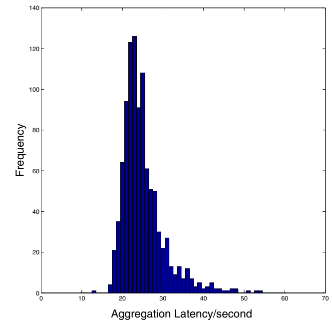


Fig. 4. Distribution of Aggregation Latency

VI. Conclusion

In our paper we defined a data aggregation model in UWSNs and proposed a Shortest Hop Tree based algorithm together with its theoretical bound. Specifically, we introduce the concept of Virtual Slot to efficiently exploit multiplexing opportunities of the time domain. The key feature of virtual slots is that they are allowed to overlap and be assigned with different slot sizes. However, note it is the layer by layer property that makes our multiplexing time utilizing scheme easy to understand and implement, thus it is hard to apply the scheme to SDA like algorithms to reduce its aggregation latency.

The idea of virtual slot can be extended to obtain more efficient time domain utilization, if we investigate more intensively on the statistics of propagation delays. At last, owning to the simplicity brought by the layer by layer property, it is probably to expand our multiplexing time utilizing scheme to realistic environment where packet loss, Doppler effect, small displacement of sensors and etc. must be considered.

Acknowledgment

This work was supported in part through National Natural Science Foundation of China (No.60803115, No.60873127, No.61073147), National Natural Science Foundation of China - Microsoft Research Asia (No.60933012).

References

[1] I. F. Akyildiz, D. Pompili and T. Melodia, "State-of-the-Art in Protocol Research for Underwater Acoustic Sensor Networks", In *ACM International Workshop on UnderWater Networks. 2006*.
[2] P.-J. Wan, K.M. Alzoubi and O. Frieder, "Distributed Construction of Connected Dominating Set in Wireless Ad Hoc Networks", In *IEEE INFOCOM 2002*.
[3] H. Shi, D. Kruger and J. V. Nickerson, "Incorporating environmental information into underwater acoustic sensor coverage estimation in estuaries", In *IEEE MILCOM 2007*, pages 1-7.
[4] M. Stojanovic, "Acoustic (underwater) Communications, Encyclopedia of Telecommunications", John Wiley and Sons, 2003.
[5] Xujin Chen, Xiaodong Hu and Jianming Zhu, "Minimum data aggregation time problem in wireless sensor networks", In *Lecture Notes in Computer Science 3794*, pages 133-142. Springer, 2005.
[6] Scott C.-H. Huang, Peng-jun Wan, Chinh T. Vu, Yingshu Li and Frances Yao, "Nearly Constant Approximation for Data Aggregation Scheduling in Wireless Sensor Networks", In *IEEE INFOCOM 2007*, pages 6-12.
[7] Bo Yu, Jianzhong Li and Yingshu Li, "Distributed Data Aggregation Scheduling in wireless sensor networks", In *IEEE INFOCOM 2009*, pages 2159-2167.
[8] Peng-jun Wan, Scott C.-H. Huang, Lixin Wang, Zhiyuan Wan and Xiaohua Jia, "Minimum-Latency Aggregation Scheduling in Multihop Wireless Networks", In *ACM Mobihoc 2009*, page 185–193.
[9] Piyush Gupta and P. R. Kumar, "The capacity of wireless networks", In *IEEE Transactions on Information Theory*, 46, pages 388–404, 2000.
[10] Jun-Hong Cui, Jiejun Kong, Mario Gerla and Shengli Zhou, "Challenges: Building Scalable and Distributed Underwater Wireless Sensor Networks (UWSNs) for Aquatic Applications", UCONN CSE Technical Report, 2005.
[11] Eduardo F. Nakamura, Antonio A. F. Loureiro, Alejandro C. Frery, "Information Fusion for Wireless Sensor Networks: Methods, Models, and Classifications", In *ACM Computing Surveys*, 39, pages 9/1–9/55, 2007.