A STATISTICAL APPROACH OF NODE ESTIMATION IN UNDERWATER WIRELESS SENSOR NETWORK

Abu Sadat Md. SAYEM$^1$      Md. Shamim ANOWER$^2$

$^1$Department of Electrical & Electronic Engineering, Rajshahi University of Engineering & Technology, Bangladesh. $^2$sayem_ruet_eee@yahoo.com, md.shamimanower@yahoo.com

Abstract: Estimating the number of active nodes in underwater wireless sensor networks is a challenging issue because a lot of factors are involved in underwater environment that create difficulties in wave propagation. The major obstacles that interfere with wave propagation in underwater network are high propagation delay, high absorption and dispersion. Although, many estimation techniques exist for node estimation in wireless sensor network but they are not efficient in underwater environment. Researchers are trying to find out more efficient method for node estimation in underwater wireless sensor network. In this paper a statistical signal processing method is proposed for node estimation in underwater wireless sensor network. Here, the nodes are considered as acoustic signal sources and their number is calculated through the cross-correlation of the acoustic signals received at three sensors placed in the network. The mean of the cross-correlation function depends on the number of signal sources and in this proposed work the mean of the cross-correlation function is used as the estimation parameter. Theoretical and simulation results are provided which reflects the validity of this cross-correlation based technique. The performance of this proposed method is evaluated by comparing the error in estimation with previous approaches.

Key words: Node, Cross-correlation function ($CCF$), Mean of cross-correlation function, Underwater acoustic sensor networks ($UASN$), Binomial probability distribution, Coefficient of variation.

1. Introduction

Nodes are deployed in underwater wireless sensor networks for a variety of applications. These applications range from research to security purposes. The most common applications are environmental monitoring, discovering natural resources, information for scientific analysis. In any type of underwater sensor network, proper network operation depends on the number of active nodes. So, estimating the number of active nodes is an important issue in any sensor network. Moreover, maintenance and localization activities need exact estimation of nodes.

Many estimation techniques exist at present to count the number of nodes. For example, a radio frequency identification (RFID) protocol specifies the algorithms for the reader and the tags so that the reader can properly collect all the tag IDs. Knowing the numbers of tags in large-scale RFID systems is possible by using the existing protocol to identify individual tags and then computing the cardinality of the system. Tag identification protocols can be divided into probabilistic [1-4], deterministic [5, 6] and hybrid [7, 8]. The protocols based techniques are similar to the estimation of the number of nodes in wireless communication networks. A node estimation technique for terrestrial sensor networks was investigated by Budianu et al. [9] – [11]. It involves an estimation based on the Good Turing (GT) estimator [12] of the missing mass. The conventional methods do not take into account the capture effect. One solution has been proposed in [13], [14] which proposed a node estimation technique taking the capture effect into account. Still it suffers from long propagation delays, high path loss in underwater acoustic network.

So, it can be said that node estimation in underwater sensor network is not same as other type of networks. In underwater environment conventional estimation techniques are not efficient. They are only applicable to communication friendly networks. Conventional methods face difficulties in underwater environment because underwater propagation characteristics [15] are suffered from high propagation delay, high absorption, and dispersion. Moreover, limited battery power, limited bandwidth, high bit error rate, short lifetime of underwater sensors create major challenges in the design of underwater acoustic networks [16], [17].

In this paper, a node estimation technique is proposed which can be more efficient method than the existing methods used in underwater network for node estimation. The proposed method is based on
the cross-correlation [18-21] of the acoustic signals received at three sensors in the network. In the proposed estimation technique the mean of cross-correlation function (CCF) is used as the estimation parameter. The transmitted signals from a number of different random signal sources (nodes) within range are received by three sensors placed in the network; the received signals are summed at each of the three sensor locations. Then, for each pair of sensors the CCFs and the mean of the CCFs are calculated. After calculating the mean of CCFs for each pair of sensors their average will be calculated. The estimation of the number of signal sources (assumed in our case the number of nodes in an underwater network) can be obtained based on the average of these mean of the CCFs. Finally, the error in estimation of the proposed technique will be compared with other techniques.

2. Formation of CCF from random signal sources

Let us, consider a 3D space where three receiving nodes are placed in a triangular form and are surrounded by N transmitting nodes as shown in Fig.1. Assuming that the transmitting nodes are the sources of white Gaussian signals and are uniformly distributed over the volume of a large sphere inside a cube, because only a sphere provides equal amounts of signals from every direction. In this case, three sensors form an equilateral triangle inside the spherical network, where the centre of the sphere lies at the centroid of that triangle as shown in Fig.1.

For the origination of CCFs in triangular sensor (TS) case, three sensors, H1, H2 and H3, and a node N1, are placed at locations (x1, y1, z1), (x2, y2, z2), (x3, y3, z3) and (x4, y4, z4), respectively (using rectangular coordinate system), somewhere inside the network as shown in Fig. 2.

Fig. 2. Underwater network with three sensors and only one node N1.

To formulate CCF, the 3D space is considered as a cube such that the dimension of the cube is equal to the diameter of the sphere and three sensors, H1, H2 and H3, and a node N1, are placed at locations (x1, y1, z1), (x2, y2, z2), (x3, y3, z3) and (x4, y4, z4), respectively (using rectangular coordinate system), somewhere within the cube. The distances between the sensors are then:

\[
\begin{align*}
\text{d}_{\text{DBS}_{12}} &= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \\
\text{d}_{\text{DBS}_{13}} &= \sqrt{(x_1 - x_3)^2 + (y_1 - y_3)^2 + (z_1 - z_3)^2} \\
\text{d}_{\text{DBS}_{23}} &= \sqrt{(x_2 - x_3)^2 + (y_2 - y_3)^2 + (z_2 - z_3)^2}
\end{align*}
\]

Here, \(d_{\text{DBS}_{12}}\) = distance between H1 and H2, \(d_{\text{DBS}_{23}}\) = distance between H2 and H3, \(d_{\text{DBS}_{13}}\) = distance between H1 and H3. These sensors are located such that, \(d_{\text{DBS}_{12}} = d_{\text{DBS}_{23}} = d_{\text{DBS}_{13}} = d_{\text{DBS}}\).

The propagation velocity is constant, which is the proposed case, the sound velocity \(S_p\), in the medium. Now, getting probe request, a node emits a very long Gaussian signal, which is recorded by the sensors with corresponding time delays. The signals in the sensors are cross-correlated, which takes the form of a delta function as it is a cross-correlation of two
white Gaussian signals where one signal essentially is a delayed copy of the other. The position of this delta in the CCF will be the distance equal to the delay difference of the signals from the centre of the CCF where the position is called a bin in this paper. This holds for all nodes and the formation of CCF for N number of nodes can be expressed as follows [22]:

Consider that node N1 emits a Gaussian signal $S_1(t)$, which is infinitely long. Then the signals received by $H_1$, $H_2$ and $H_3$ are, respectively

$$S_{r_{11}}(t) = a_{11}S_1(t - \tau_{11}) \quad (1)$$

$$S_{r_{12}}(t) = a_{12}S_1(t - \tau_{12}) \quad (2)$$

$$S_{r_{13}}(t) = a_{13}S_1(t - \tau_{13}) \quad (3)$$

where, $a_{11}$, $a_{12}$ and $a_{13}$ are the respective attenuations due to the absorption and dispersion present in the medium, and $\tau_{11} = \frac{d_{11}}{s_p}$, $\tau_{12} = \frac{d_{12}}{s_p}$ and $\tau_{13} = \frac{d_{13}}{s_p}$ are the respective time delays for the signal to reach the sensors.

Thus for a dense network having N nodes, if the transmitted signals from the nodes are denoted as $S_1(t)$, $S_2(t)$, $S_3(t)$, ..., $S_N(t)$ respectively, the corresponding delays to reach $H_1$ are denoted as $\tau_{11}$, $\tau_{21}$, ..., $\tau_{N1}$, and the corresponding attenuations are as $a_{11}$, $a_{21}$, ..., $a_{N1}$, the composite signal received by $H_1$ can be expressed as:

$$S_{r_{c1}}(t) = \sum_{j=1}^{N} a_{1j}S_j(t - \tau_{1j}) \quad (4)$$

Similarly, if the transmitted signals from N nodes are denoted as $S_1(t)$, $S_2(t)$, ..., $S_N(t)$ respectively, the corresponding delays to reach the sensors $H_2$ and $H_3$ are denoted as $\tau_{12}$, $\tau_{22}$, $\tau_{N2}$, and $\tau_{13}$, $\tau_{23}$, ..., $\tau_{N3}$, respectively, and the corresponding attenuations are as $a_{12}$, $a_{22}$, ..., $a_{N2}$, and $a_{13}$, $a_{23}$, ..., $a_{N3}$, respectively, the composite signals received by the sensors $H_2$ and $H_3$ can be expressed as:

$$S_{r_{c2}}(t) = \sum_{j=1}^{N} a_{2j}S_j(t - \tau_{2j}) \quad (5)$$

and

$$S_{r_{c3}}(t) = \sum_{j=1}^{N} a_{3j}S_j(t - \tau_{3j}) \quad (6)$$

respectively.

Then, the CCFs between the signals at ($H_1$, $H_2$), ($H_2$, $H_3$) and ($H_3$, $H_1$) are, respectively:

$$C_{12}(\tau) = \int_{-\infty}^{+\infty} S_{r_{c1}}(t) S_{r_{c2}}(t - \tau) dt \quad (7)$$

$$C_{23}(\tau) = \int_{-\infty}^{+\infty} S_{r_{c2}}(t) S_{r_{c3}}(t - \tau) dt \quad (8)$$

$$C_{31}(\tau) = \int_{-\infty}^{+\infty} S_{r_{c3}}(t) S_{r_{c1}}(t - \tau) dt \quad (9)$$

which take the form of a series of delta functions [22]. Here, $\tau = d_{dls} / s_p$, is the time shift in cross-correlation. One such CCF obtained with $N = 1000$ nodes is shown in Fig. 2. Bins, $b$ in the CCF (as shown in Fig. 3) is defined as a place occupied by a delta inside a space of a width twice the distance between sensors and that place is determined by the delay difference of the signal coming to the sensors. The deltas of equal delay differences are placed in that particular bin.

![Fig. 3. Bins, $b$ in the cross-correlation process](image)

### 3. Counting the number of nodes from the mean of CCF

It is known that the cross-correlation function produced from white Gaussian signals follows the binomial probability distribution [22] in which the parameters are the number of nodes, $N$, and the number of bins, $b$ [22]. Then the expected value, i.e. the mean, $m$ of the CCF is defined as [22]

$$m = \frac{N}{b} \quad (10)$$

where $b$ is the number of bins in the cross-correlation process and is obtained from the experimental setup with sampling rate, $s_r$, distance between sensors, $d_{dls}$, and speed of propagation, $s_p$ as [22], [23]:

$$b = \frac{2 \times d_{dls} \times s_r}{s_p} - 1 \quad (11)$$
Thus the estimation of \( N \) is obtained from equation (10) as:

\[
N = b \times m
\]  
(12)

This is the relationship between the number of nodes, \( N \), and the mean, \( m \), of the CCF. Since we know \( b \) and can measure \( m \) from the CCF, we can readily determine the number of nodes, \( N \). Now, the theory for node estimation from the mean CCF for three sensor case are discussed below:

The estimation parameter, \( m_{\text{average}}^{3\text{CCF}} \), is obtained by taking mean, \( m_{12}, m_{23}, \) and \( m_{31} \), from three CCFs. \( m_{\text{average}}^{3\text{CCF}} \) can be expressed as:

\[
m_{\text{average}}^{3\text{CCF}} = \frac{N}{b_{12}} + \frac{N}{b_{23}} + \frac{N}{b_{31}}
\]
(13)

Here, \( b_{12} = b_{23} = b_{31} = b \); so, it is found that

\[
m_{\text{average}}^{3\text{CCF}} = \frac{m_{12} + m_{23} + m_{31}}{b}
\]
(14)

From \( m_{\text{average}}^{3\text{CCF}} \), the number of nodes, \( N \), can be calculated easily.

4. Simulation Result

Now, a setup will be employed to perform the simulations to verify the theoretical results by those from the simulations. Simulations have been performed by Matlab programming. For simulation environment it is assumed that three sensors are placed in triangular shape somewhere in the middle of a sphere inside a cube such that the diameter of the sphere is equal to the dimension of the cube, and the sphere is filled with a number of uniformly distributed nodes that emits white Gaussian signals. The other parameters are radius of the sphere is 2000m; \( N=1,10, 20,...,100; \) signal length is \( 10^6 \) samples; signal propagation speed is 1500m/s; absorption coefficient is 1; dispersion factor is 0, equal received powers from all nodes, no background noise and sampling rate \( S_R = 180 \text{ kSa/s} \). The signals (responses to probe requests from the sensors or autonomous) emitted from the nodes are collected by the sensors. By cross-correlating these three signals at the sensors, three set of CCF are obtained. Using these CCF, the estimation tool, i.e., the average of the mean of the CCF is calculated. Simulation results for different bin number are shown in the following Figs.
Fig. 4 to Fig. 6 show that the theoretical and corresponding simulated results for the estimation of the number of nodes in a network in terms of the estimation parameter, $m_{CCF}$, which show that the simulations match the theory properly and is the indication of effectiveness of the process. The solid lines indicate the theoretical results and the circles the corresponding simulated results. The variations of $b$ in the three different Figs. are as a result of varying $d_{DBS}$ (sampling rate and propagation speed is kept constant). The distances between the sensors are: 0.125m in Fig. 4, 0.25m in Fig. 5 and 0.5m in Fig. 6.

However, the above results indicate that the process is applicable for node estimation in underwater environment. At the same time, it is clear that the number of bins, $b$ affects the estimation parameter and it can be seen that the value of the estimation parameter is lower in case of higher $b$ and vice-versa and simulated results more perfectly matched with the theoretical results. By taking more number of bins better result can be achieved. Number of bin can be varied by varying sampling rate, propagation velocity or the distance between sensors as expressed in equation (11).

Now, another approach will be taken, the sampling rate will be doubled (360kHz/s) and the process will be repeated. A comparison will be observed for the estimation process for the same number of bins as before.
From Fig. 7 to Fig. 9 it can be observed that improvement in result occurs with the increase in number of bins as previous sampling rate.

Now, the results will be shown for the estimated number of nodes, \( N(\text{estimated}) \) with respect to exact number of nodes.

![Comparison of theoretical and simulated number of estimated nodes](image)

Fig. 10 shows the comparison of theoretical and simulated number of estimated nodes (for bin number 119). In this Fig., the solid line indicates the theoretical result and the circles the corresponding simulated results. From Fig. 10, it is clear that, the theoretical and simulated results are very close to each other, which indicates the validity of the proposed approach.

### 5. Analysis of error in estimation

There is always an error presents in every simulation and the quantity of error represents the performance of the estimation. The error of the estimation can be measured in various ways. Numerically, estimation error can be represented in different ways: such as-(i) as a true error, or (ii) as a statistical error. A true error is preferable when the parameter used in the experiment is not random, i.e., it gives a fixed estimation every time for a particular setup. Whereas, in an experiment with random numbers, as the estimated values vary from time to time for a particular setup, thereby indicating a certain statistical property, it is better to represent the error statistically. As the proposed cross-correlation is a statistical technique, the statistical error, the coefficient of variation (CV), is used as its error in estimation in order to fully assess the accuracy of the proposed estimation techniques. To obtain a simulated CV of estimation, a simulation process is run 1000 times for a particular \( N \) and \( b \). From these 1000 values of estimated \( \hat{N} \), the standard deviation and mean of estimation and, thus, the CV, are obtained. In this case firstly, the mean of the CCF from 100 iterations, and then the estimated \( \hat{N} \) using the expression of \( N \) related to this mean, are obtained. Secondly, to obtain the CV, the same process is continued 1000 times without any change in parameters and the values of all estimated \( \hat{N} \) are recorded. Finally, the CV for one iteration is obtained from the ratios of the standard deviation to the mean of those values as [24]:

\[
CV = \frac{\sigma(N)}{\mu(N)}
\]

Now, if \( u \) iteration is used, the standard deviation and, thus, the CV, are reduced to \( \frac{1}{\sqrt{u}} \) so that the CV after the \( u^{th} \) iteration is

\[
CV = \frac{1}{\sqrt{u}} \left( \frac{\sigma(N)}{\mu(N)} \right)
\]

### 6. Comparison of error with previous estimation techniques

Now, the error involved in estimation process of the proposed technique will be compared with those of previous techniques: the probabilistic framed slotted ALOHA (PFSA) [1]; the Good-Turing (GT) [10] estimator protocol, DIIPUC [13, 14]; two sensors cross correlation technique based on the ratio of mean and standard deviation of CCF [24] and two sensors cross correlation technique based on the mean of CCF [25]. The CVs are compared keeping the estimation time fixed.

In the above Fig.11 CCF: MEAN (3 sensors) is the proposed method, CCF: MEAN (2 sensors) is the two sensors cross correlation technique [25] which is the latest work before the proposed method which used two sensors and the mean of CCF was the estimation parameter, CCF: RATIO is two sensors cross correlation technique [24] and rest three are conventional protocol-based techniques.
In the above comparison it is considered a very long fixed signal length, $N_s$, of 158093 samples, sampling rate 390000 HZ, signal propagation speed 1500m/s, bin number119 and $d_{BS}=0.25$m. For conventional protocol-based techniques the considered values of the parameters are: First frame size, $F_1=512$; the maximum transmission range of the probing node is $R_t=2000$m; the number of bits per packet is $B_n= 112$ bit/packet. The bit rate of the channel, $B_c= 15$kbps considering 15kHz bandwidth and BPSK modulation technique, number of packets per slot, $\rho =1$for DIIPUC, $\rho$ is 4 is GT, $\rho$ is 1.59 for PFSA and estimation time is $40.5367$ seconds.

Corresponding to CV it can be concluded that the proposed approach gives better accuracy than previous methods. Although Good Turing (GT) [11] method is better for fewer numbers of nodes the proposed technique is better for dense network which is usual case for underwater network.

7. Conclusion

In this paper a theoretical models and corresponding techniques of node estimation in underwater wireless sensor network has been proposed. Later a simulation has been performed assuming an environment similar to a dense underwater wireless sensor network. From the simulation results it has been observed that the proposed method is efficiently applicable for node estimation in dense underwater wireless sensor network. Also, error in estimation is calculated and compared with previous techniques which indicate that the proposed approach is better than other techniques in underwater environment. In this paper no practical work has been performed which can be performed in ocean wireless sensor network in further research. In addition, locating the exact physical positions of the nodes can be analyzed in future research.

References


