Source Traffic Modelling in WSN for Acoustic Sensing in Reverberant Environment

Nenad B. Cetic[1], Miroslav V. Popovic[1], Miodrag M. Djukic[1], Jelena V. Kovacevic[1]

[1]Department of Computer Engineering and Communications, Faculty of Technical Sciences,
University of Novi Sad,
Trg Dositeja Obradovica 6, 21000 Novi Sad, Serbia
nenad.cetic@rt-rk.com

[1]Abstract—**Recent research efforts show that Wireless Sensor Networks (WSN) for acoustic sensing have the potential for ubiquitous sensing. The key challenge in delivering such WSN architecture is the trade-off between the cost of wireless sensors and acoustic analysis complexity. In this paper, we are proposing a source traffic modelling considering a system in which sensors are connected to a cloud based processing service. Acoustic sensing in practice generates great amount of traffic, and therefore a new model of sensing in a reverberant environment is necessary for better understanding of traffic nature and its volume. To lower the sensor cost, we have considered low-complexity acoustic event detection on the sensor side for preventing unnecessary traffic. The proposed acoustic sensor acts like a noise gate. Sound propagation in reverberant environment is modelled using the modified Image-Source Modelling (ISM) method. The source traffic model is validated in Matlab simulation and in real reverberant environment. Results show an example of bandwidth used by a simple acoustic sensor.**

Index Terms—**Traffic model, wireless sensor networks, acoustic sensing, reverberant environment, room impulse response, image-source method, noise gate.**

I. INTRODUCTION

The main paradigm of the internet of things is providing services using low-resource devices connected over the conventional internet network. In this paper, we are considering an approach in which sensors are connected to a cloud-based processing service. Sensors connected to powerful processing units can provide many techniques of sensor data fusion. Traditionally microphone grids are used as a powerful tool for extracting acoustic data, source localization or de-noising. Recent research shows a good potential of using module-clouds [1] and streaming of audio [2], [4]. The system that we have in consideration is deployed using the existing internet infrastructure. Obviously, a bottleneck is the link to the cloud-based processing. This leads to the problem of having conventional network implementation which uses routers based on FIFO buffers that cannot assure quality of service. The usual "best effort" service cannot guarantee bandwidth, delay, or prevent a packet loss.

A protocol such as Real-time Transport Protocol (RTP) includes an ordinal number and timestamp for each packet. WSN needs to be able to synchronize clocks for every sensor so that timestamps are reasonably correlated. IEEE 1588 packet-based timing mechanism can be used on an embedded device to synchronize clocks with the resolution of under a microsecond [5]. This can provide good temporal fidelity on the receiver side, which is essential for the analysis typical of microphone array systems. Still, what can be an impediment for the real-time system performance is the delay introduced by digital signal encoder on the acoustic sensor node and the time for the data transferring through conventional internet network.

The existing research of source traffic modelling is not applicable to WSN for acoustic sensing. Data traffic model in WSN for intrusion detection is considered in [6]. Reference [7] shows the traffic of the medical sensors data that is modelled on the basis of an empirical data set. Both models are important contributions in this field but they cannot be used for the WSN that we have in consideration. The approach described in [8] is to some extent applicable to WSN for acoustic sensing since the signal nature in the system for target tracking is the same as the acoustic signal. However, the nature of the sound propagation in reverberant environment needs to be taken into account. In order to analyse properties of WSN source traffic, before the described system is deployed, a new model has to be derived.

II. SOUND PROPAGATION MODEL

For modelling sound propagation in reverberant environment Allen and Berkley proposed a very accurate method called Image-Source Modelling (ISM) [9]. The ISM operates most simply with rectangular rooms, but it is not restricted only to such enclosures. In theory, any shape of an enclosed space can be modelled with ISM. A recent paper demonstrates object tracking in non-convex rooms using ISM in reverse direction [10]. In this paper we have only modelled rectangular enclosures, considering the fact that majority of living space rooms can be approximated as rectangular. The basic idea is to calculate impulse response function between any two points in a small reverberant room taking into account room dimensions, shape, wall reflection properties, etc. When the room impulse response (RIR) is generated it can be convoluted to any desired input signal

simulating reverberation effects. The original ISM is computationally intensive. In a more recent paper, Lehmann and Johansson [11] proposed an improved method in which energy decay estimation is used.

Here we are giving a minimal subset of ISM equations, necessary to express and understand simulation parameters used in our experimental work. Cartesian coordinate system is used as follows:

$$
\begin{cases}
p_{source} = \begin{bmatrix} x_{source} \\ y_{source} \\ z_{source} \end{bmatrix}, \\
p_{sensor} = \begin{bmatrix} x_{sensor} \\ y_{sensor} \\ z_{sensor} \end{bmatrix}, \\
d_{room} = \begin{bmatrix} D_x \\ D_y \\ D_z \end{bmatrix},
\end{cases}
\tag{1}
$$

where p_{source} is the position of a sound source, p_{sensor} is the position of a point of interest and d_{room} is the dimension of the room. The simulator calculates RIR in time domain $h(t)$ for every point of interest using

$$
h(t) = \sum_{u_{sum}=0}^{1} \sum_{l_{sum}=-\infty}^{\infty} A(u_{sum}, l_{sum}) \times u(t - \ddagger(u_{sum}, l_{sum})), \tag{2}
$$

where $(\cdot)$ denotes the Dirac impulse function, and the $u_{sum}(u, v, w)$ and $l_{sum}(l, m, n)$ represent a triple sum over each of the triplet's internal indices respectively. Triplet's (u, v, w) and (l, m, n) are parameters for indexing the image source parameters in all dimensions. The attenuation factor $A(u_{sum}, l_{sum})$ and the time delay (u_{sum}, l_{sum}) in (2) are defined as follows:

$$
A(u_{sum}, l_{sum}) = \frac{s_{x,1}^{|l-u|} s_{x,2}^{|l|} s_{y,1}^{|n-v|} s_{y,2}^{|m|} s_{z,1}^{|n-w|} s_{z,2}^{|n|}}{4f \times d(u_{sum}, l_{sum})}, \tag{3}
$$

$$
\ddagger(u_{sum}, l_{sum}) = d(u_{sum}, l_{sum}) / c. \tag{4}
$$

As shown in (3) we get transformation function for simulating reverberation effect due to reflection from enclosure and delay. Reflection is modelled by reflection coefficient (3). Delay (4) is due to sound propagation at the velocity of sound c. Coefficients are calculated using absorption coefficients

$$
r = 1 - s^2. \tag{5}
$$

The distance $d(u_{sum}, l_{sum})$ from the image source to the sensor node is calculated using Euclidean norm of multiplication between diagonal matrix $diag(\cdot)$ with parameters (u, v, w) and (l, m, n), and p_{source} and d_{room}, respectively

$$
d(u_{sum}, l_{sum}) = \left\| \begin{array}{c} diag(2u-1, 2v-1, 2w-1) \times p_{source} + \\ p_{sensor} - diag(2l, 2m, 2n) \times d_{room} \end{array} \right\|, \tag{6}
$$

In the simulation we use RIR H( ) computed for frequency domain as

$$
H(\check{S}) = \sum_{u_{sum}=0}^{1} \sum_{l_{sum}=-\infty}^{\infty} A(u_{sum}, l_{sum}) \times \\ \times \exp(-j\check{S}\ddagger(u_{sum}, l_{sum})), \tag{7}
$$

where is the frequency variable. Signal calculation $y_{sensor,i}(t)$ for the sensor i starts with discrete Fourier transform $F(y_{source}(t))$. Then, convolution in frequency domain is performed with the transformation function $H_i(\ )$. The signal is then transformed back to time domain with inverse Fourier transform $F^{-1}\{Y(\ )\}$

$$
y_{sensor,i}(t) = F^{-1}\left\{ F\left\{ y_{source}(t) \right\} \otimes H_i(\check{S}) \right\}. \tag{8}
$$

For the sake of brevity, a detailed derivation of absorption coefficient is omitted. The derivation example is given in Lehmann and Johansson's paper [11].

III. SENSOR NODE MODEL

When WSN described in this paper is deployed, we are assuming that sensor nodes use low-cost microphones, for practical and economic reasons. Therefore, it is expected that sound fidelity cannot be high. This fact is introducing a limited dynamic range of sensing capabilities. Each sensor node has to perform a self-adjusting to its physical characteristics. Sensing is performed by continuously recording sound locally in the node. Sensor node should send data only in the presence of acoustic events. Also, it is an imperative that WSN is energy efficient. Therefore, a simple acoustic sensor with a noise gate (SASwNG) is proposed. We use digital signal processing (DSP) algorithm similar to [12]. In our case, a less computationally intensive foreground to background separation is performed so it is less accurate. The main idea of our modelling and simulation is to evaluate the performance of a simple noise gate in reverberant environment.

The recorded signal $y(t)$ is represented in the time domain as

$$
y(t) = s(t) + n(t), \tag{9}
$$

where $s(t)$ is a signal of interest and $n(t)$ represents stationary noise. The algorithm for signal power estimation $\widehat{P}(y(t), k, N)$ of a discrete signal $P(y(t), k, N)$ for each N sample block, starts from the block at the sample k by calculating the mean squared value of the signal

$$
P(y(t), k, N) \approx \widehat{P}(y(t), k, N) = \frac{1}{N} \sum_{t=k}^{k+N} \left| y(t)^2 \right|. \tag{10}
$$

When powered up, SASwNG has to be calibrated. Sensor calibration is performed on the first recorded samples M $(t = 0; k = 0)$. As a result of calibration, we have the power of noise estimated as

$$
P(n(t)) \approx \widehat{P}(n(t), 0, M) = \widehat{P}(y(t), 0, M). \tag{11}
$$

Calibration is performed on the assumption that the signal $s(t)$ is zero for M samples. This measure is considered as power of the noise floor containing both the imperfection of a sensor and the environmental noise. After calibration, signal-to-noise ratio (SNR) estimation of the signal is calculated as

$$SNR(s(t),k,N) = 10\log_{10}\left(\frac{\hat{P}(y(t),k,N)}{\hat{P}(n(t),0,M)}\right). \qquad (12)$$

The SNR estimation is compared to $SNR_{threshold}$ parameter for noise gate algorithm. The algorithm is operating on N samples blocks replacing blocks without significant acoustical event with a block of silence. Here we are introducing new parameters for desired sensor hysteresis. Attack time T_{attack} is the time for noise gate to start releasing packets. Similarly, release time $T_{relaease}$ is the time for noise gate to stop releasing packets. In practice, T_{attack} is set to the value that is by magnitude smaller than $T_{relaease}$. When T_{attack} and $T_{relaease}$ are set to zero, the simplified algorithm looks like this

$$y(t) = \begin{cases} \forall y(t), k < t < kN, \\ y(t), SNR(y(t),k,N) > SNR_{threshold}, \\ -\infty dB, otherwise. \end{cases} \qquad (13)$$

The equation (8) is producing a signal modified by sound propagation. When that signal is used as input to (13) we get noise gated signal. This is in essence our proposed source traffic model.

We assume that the background noise level varies over time, more slowly than the sounds of interest, and that no sharp frequency peaks are present in the background noise. SASwNG is not sending block of silence. Each block is sent in one network datagram that consists of the header and the payload. The payload consists of N samples of audio data. In the case of RTP over UDP/IPv4, the header size is 40 octets. To keep SASwNG as simple as possible we assume that no compression is performed on audio data as suggested in [13]. RTP packet contains raw 16bit PCM data. Figure 1 illustrates an example of datagram generation.
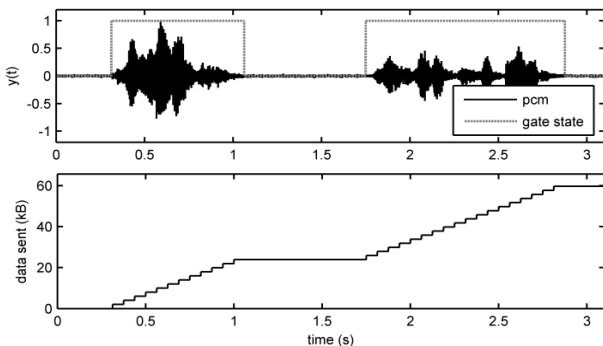

Fig. 1. SASwNG source traffic generation for voice recorded sequence. Gate state shown as (0 – gate closed; 1 - gate open). Parameters: M = 5000 samples; N = 1000 samples.

Different aspects of the model are shown in Table I. It is the comparison of our paper to related papers in references.

Here, we try to provide a clear overview of relations to the existing papers from references. It is shown that compression of data is usually performed before sending. To provide the exact data rates calculation we are not considering any data compression.

TABLE I. COMPARISON WITH RELATED WORK.

Aspect	Our work	Related work
ISM	Eq. (1)–(8)	Existing ISM [9], [11]
ISM implementation	Processing data chunks	Batch processing. Loading all data in memory.[14]
Foreground vs. background separation	Low complexity Eq. (9)–(13); Based on energy estimation	Acoustic monitoring [12] Great complexity; Based on gamma-tone filter-bank
Packets	Raw PCM data. No compression.	Compressed data[1], [2], [4] Metadata packets[3] No compression indicated [13]

IV. EXPERIMENTAL RESULTS

The acoustic simulation of our room is based on Lehmann's work available at [14]. Our main contribution to the ISM simulation consists of reworking implementation of the main loops in the simulation program to enable processing of a large audio data set. The original implementation loads the whole input audio file. In addition the input audio file can have multiple source signals and calculated ISM in a batch process from loading input to the production of all output signals. This caused "out of memory" problem so we had to reorganize processing to operate on the data chunks.

After that, ISM validation is done. Since SASwNG uses signal power estimation, energy decay is measured in a simulated room. The signal source is placed at the stationary point. Sensors are placed to cover the whole area of the room. Validation is done using various signals at the source by calculating the root mean square (RMS) value. RMS value is expressed in decibels

$$RMS_{dB} = 20\log_{10}(RMS(y(t)), \qquad (14)$$

where $0\ dB$ represents full scale amplitude. Figure 2(a) shows the case without reverberation when only direct sound path propagation is calculated. The generated white noise is used as an input signal. We can see that signal energy decays with d^2. In Fig. 2(b) the same input signal is used but now SASwNG is placed in the simulated reverberant environment. We can see that the signal energy decays more slowly due to reverberation but with a uniform distribution. The next case in Fig. 2(c) shows superposition of waves from direct path with reflected waves. The choice of sine wave at 128 Hz provides enough resolution to see wave pattern in energy distribution. When 1 kHz sine wave is used complex energy decay pattern is produced as Fig. 2(d) demonstrates. Both patterns on Fig. 2(c) and Fig. 2(d) are caused by additive and subtractive wave interference depending on sensor position. Simulation parameters are listed in Table II. T_{60} is a simulation parameter that specifies the desired decay time. It is the time needed for the signal to drop by 60dB.
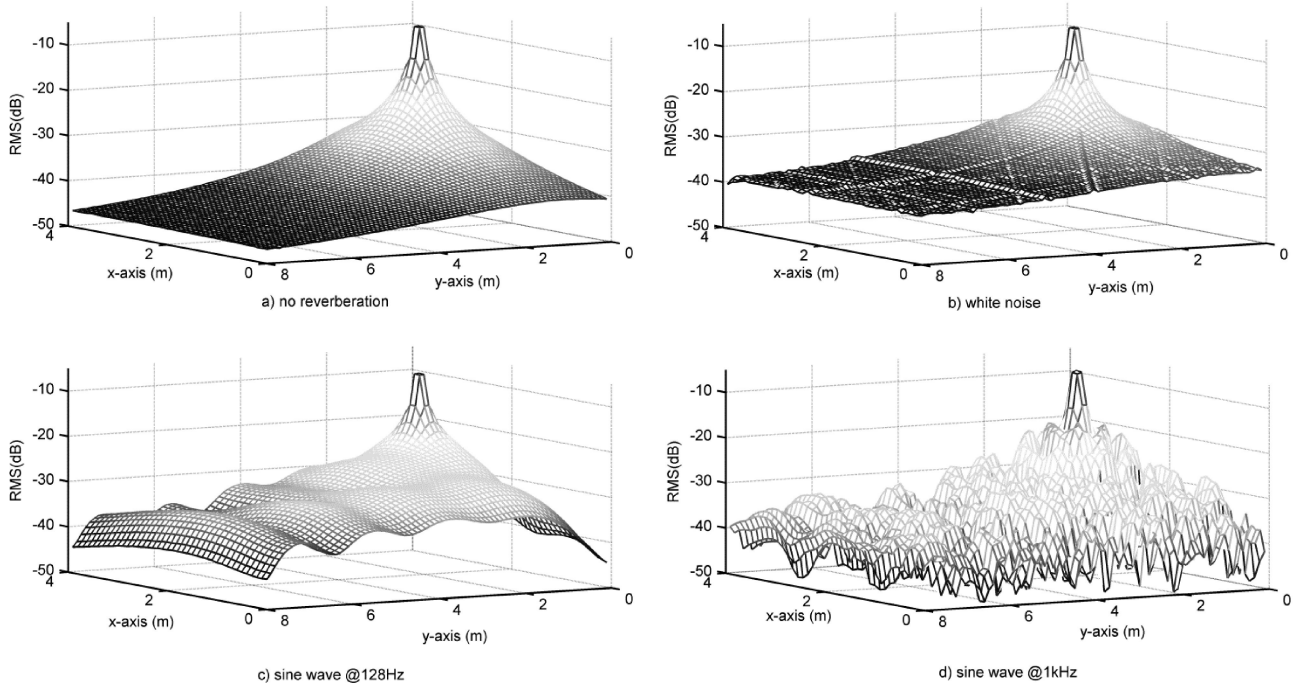
Fig. 2. Sound propagation simulation using ISM. Different input signal used for physical verification: (a) white noise in anechoic environment, (b) white noise in reverberant environment, (c) 128 Hz sine wave, (d) 1 kHz sine wave. All input signals scaled at RMS = -9.01dB.

TABLE II. ISM SIMULATION PARAMETERS.

Symbol	Parameter description	Parameter value
d_{room}	room dimension	[4 m , 8 m, 2.3 m]^T
p_{source}	source position	[3 m, 1 m, 1.3m]^T
	number of sensors	40×80 = 3200 sensors
$p_{sensor,k,j}$	sensor positions	k0.1 m, j0.1 m, 1.3
	absorption coefficients	[0.82 0.81 0.71 0.69 0.37 0.42]
T_{60}	decay time	200 ms

Next, SASwNG performance verification is done using sequences of recorded voice. Authors of [15] propose high sampling frequency such as 44.1 kHz as a requirement for source localization techniques. Our data set consists of voice recordings at 16 kHz, but simulation is not limited to that frequency. Data rates are measured for SASwNG processing of the original record and the signal produced by ISM. For this experiment a sensor is placed on a different side of the room. Simulation parameters are given in Table III.

TABLE III. DATA RATE SIMULATION PARAMETERS.

Symbol	Parameter description	Parameter value
d_{room}	room dimension	[5.3 m , 2.5 m, 2.67 m]^T
p_{source}	source position	[5.0 m, 2.2 m, 1.3m]^T
p_{sensor}	sensor positions	[0.2 m, 0.2 m, 1.3m]^T
	absorption coefficients	[0.13 0.13 0.13 0.13 0.13 0.13]
T_{60}	decay time	899.5 ms
F_s	sampling rate	16000 Hz
Q	bits per sample	16
M	calibration block size	5000 samples
N	estimation block size	from 100 to 5000 samples
H	header size	40 octets = 320 bits
	signal duration ratio	0.45406915

For different block sizes *N(samples), an* average data rate is calculated. Theoretical maximum data rate is produced if we take a naïve approach when audio is streamed continuously, and all packets are sent. In general, traffic data rates $R(F_s,N,)$ produced by SASwNG are given as

$$\frac{F_s}{N}(H + NQ)\} \le R(\text{Fs,N,}) \le \frac{F_s}{N}(H + NQ), \quad (15)$$

where F_s is a sample rate, H is header size in bits, Q is number of quantization bits and is the ratio between the signal of interest duration (foreground signal) and total recorded time. Factor is derived by hand-annotation of sequences duration in which a voice is present for each recorded sequence. Maximum data rate multiplied by ratio gives a minimum theoretical data rate necessary.

To validate our findings from the simulation we have conducted experiment in the lab. An audio recording was processed by SASwNG algorithm and data rates were calculated like in the previous experiment. Measurement of absorption coefficients is a task that needs to be done with great care if we want to have a simulation that resembles the real reverberant environment. This is not done in this paper and we will consider it in the future. The decay time that match real reverberation environment is used and uniformly distributed sound absorption is assumed. Current lab results confirm our presumption once again and show how reverberation affects data rates. The results of the experiments are depicted in Fig. 3.

It is noticeable that data rate increases for higher value of N. This is due to misalignment of voice segments and processing block boundaries. In principle, this could be overcome by moving the processing window sample by sample; however, this would add significant amount of processing which would neglect the simplicity of SASwNG design. In conducted experiments in reverberant environment, SASwNG reduced traffic up to 43.78 % when compared to theoretical maximum. Experiments show that the effects of reverberation are lowering bandwidth savings.
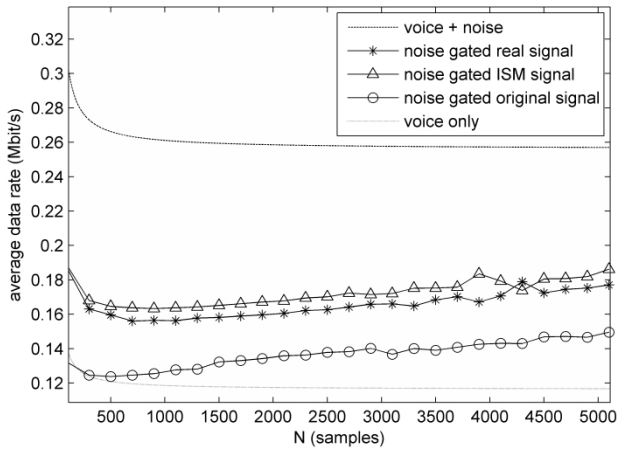
Fig. 3. Average data rates for different block sizes – N(samples).

Also, we have to state that high performance microphone arrays in indoor environment [16] will always perform better than WSN. What we have considered here is a low cost sensor that could be easily installed in a building of interest using the existing network. This approach is an approach advocated by IoT paradigm. SASwNG proposed in this paper focuses on the design aspect of source traffic modelling if sensor processing power is very low. The simple NG described here has a small memory footprint and leaves enough processing power for sensor synchronization like the one used in [17] or some other additional work in WSN.

V. CONCLUSIONS

The source traffic model for acoustic sensing in reverberant environment is presented. Traffic modelling is an important part of a network design. A good traffic model enables analysis and simulation of WSN before the actual employment is done. This can lower the cost and improve WSN efficiency. The presented source traffic model can be used for exploring traffic volume in various reverberant environments, such as small rooms and the like.

The algorithm is compared with the naïve case when the sensor generates traffic continuously. Experimental results confirm that simple separation of foreground versus background signal can save significant bandwidth. Validation of data is done with an audio recording in a reverberant room. When SASwNG is placed farther from the signal source, it is clear that it produces more data due to distortion caused by reverberation. It is true that the described WSN is indeed more suitable for algorithms such as energy-based acoustic source localization. Algorithms that rely on the difference in arrival or phase information demand predictable jitter which WSN usually cannot provide.

REFERENCES

[1] N. Nirjon, R. F. Dickerson, P. Asare, Q. Li, D. Hong, J. A. Stankovic, P. Hu, G. Shen, X. Jiang, "Auditeur: a mobile-cloud service platform for acoustic event detection on smartphones", in *Proc. 11th Annual Int. Conf. Mobile Systems, Applications, and Services (MobiSys 2013),* Taipei, Taiwan, 2013, pp. 403–416. [Online]. Available: http://dx.doi.org/10.1145/2462456.2464446

[2] C. Pham, P. Cousin, "Streaming the sound of smart cities: experimentations on the SmartSantander test-bed", in *Green Computing and Communications (GreenCom), IEEE and Internet of Things (iThings/CPSCom), IEEE Int. Conf. and IEEE Cyber, Physical and Social Computing*, Beijing, China, 2013. [Online]. Available: http://dx.doi.org/10.1109/GreenCom-iThings-CPSCom.2013.114

[3] D. Hollosi, G. Nagy, R. Rodigast, S. Goetze, P. Cousin, "Enhancing wireless sensor networks with acoustic sensing technology: use cases, applications & experiments", *IEEE Int. Conf. Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*, 2013, Beijing. [Online] Available: http://dx.doi.org/10.1109/GreenCom-iThings-CPSCom.2013.75

[4] A. Carer, C. Pham, P. Cousin, "Real-time on-demand multi-hop audio streaming with low-resource sensor motes", *IEEE 39th Conf. Local Computer Networks Workshops (LCN Workshops),* 2014, Edmonton, Canada. [Online] Available: http://dx.doi.org/10.1109/LCNW.2014.6927700

[5] D. May, "The XMOS architecture and XS1 chips", *Micro, IEEE*, vol. 2, no. 6, pp. 28–37, 2012. [Online] Available: http://dx.doi.org/10.1109/MM.2012.87

[6] I. Demirkol, F. Alagoz, H. Delic̦, C. Ersoy, "Wireless sensor networks for intrusion detection: packet traffic modeling", *Communications Letters, IEEE*, vol. 10, no. 1, pp. 22–24, 2006. [Online] Available: http://dx.doi.org/10.1109/LCOMM.2006.1576557

[7] G. G. Messier, I. G. Finvers, "Traffic models for medical wireless sensor networks", *Communications Letters, IEEE*, vol. 11, no. 1, pp. 13–15, 2007. [Online] Available: http://dx.doi.org/10.1109/LCOMM.2007.061291

[8] Q. Wang, T. Zhang, "Source traffic modeling in wireless sensor networks for target tracking", in *(PE-WASUN 2008)*, Vancouver, British Columbia, Canada, 2008. [Online] Available: http://dx.doi.org/10.1145/1454609.1454629

[9] J. Allen, D. Berkley, "Image method for efficiently simulating small-room acoustic", *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979. [Online] Available: http://dx.doi.org/10.1121/1.382599

[10] O. Ocal, I. Dokmanic, M. Vetterli, "Source localization and tracking in non-convex rooms", *39th Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP 2014)*, Florence, Italy, [Online] Available: http://dx.doi.org/10.1109/ICASSP.2014.6853833

[11] E. A. Lehmann, A. M. Johansson, "Prediction of energy decay in room impulse responses simulated with an image-source model", *The Journal of the Acoustical Society of America*, vol. 124, no. 1, pp. 269–277, 2008. [Online] Available: http://dx.doi.org/10.1121/1.2936367

[12] S. Goetze, J. Schroder, S. Gerlach, D. Hollosi, J.-E. Appell, "Acoustic monitoring and localization for social care", *Journal of Computing Science and Engineering*, vol. 6, no. 1, pp. 40–50, 2012. [Online] Available: http://dx.doi.org/10.5626/JCSE.2012.6.1.40

[13] D. Zordan, B. Martinez, I. Vilajosana, M. Rossi, "To compress or not to compress: processing vs transmission tradeoffs for energy constrained sensor networking", Technical Report. Department of Information Engineering, University of Padova, Padova, Italy 2012. [Online] Available: http://arxiv.org/abs/1206.2129

[14] E. A. Lehmann, "Fast image-source method: Matlab code", 2013. [Online] Available: http://www.eric-lehmann.com

[15] O. M. Bouzid, G. Y. Tian, J. Neasham, B. Sharif, "Investigation of sampling frequency requirements for acoustic source localisation using wireless sensor networks", *Applied Acoustics,* vol. 74, no. 2, pp. 269–274, 2013. [Online] Available: http://dx.doi.org/10.1016/j.apacoust.2010.12.013

[16] H. Atmoko, D. C. Tan, G. Y. Tian, B. Fazenda, "Accurate sound source localization in a reverberant environment using multiple acoustic sensors", *Measurement Science and Technology,* vol. 19, no. 2, 2008. [Online] Available: http://dx.doi.org/10.1088/0957-0233/19/2/024003

[17] B. Kelly, D. Hollosi, P. Cousin, S. Leal, B. Iglar, A. Cavallaro, "Application of acousting sensing technology for improving building energy effiviency", *Procedia Computer Science,* vol. 32, pp. 661–664, 2014. [Online] Available: http://dx.doi.org/10.1016/j.procs.2014.05.474