PRIVACY-PRESERVING DISTRIBUTED SPEECH ENHANCEMENT FOR WIRELESS SENSOR NETWORKS BY PROCESSING IN THE ENCRYPTED DOMAIN

Richard C. Hendriks and Zekeriya Erkin

Delft University of Technology, The Netherlands
{R.C.Hendriks, Z.Erkin}@tudelft.nl

Timo Gerkmann

University of Oldenburg, Germany
timo.gerkmann@uni-oldenburg.de

ABSTRACT

To improve speech communication in noisy and reverberant environments, an increased interest is shown to develop algorithms that make efficiently use of acoustic wireless sensor networks (WSNs). The processors and sensors forming these WSNs can be owned by multiple users. Sending private data across such a WSN can lead to severe privacy and security issues and may limit its acceptance. Using the advantages of WSNs, while guaranteeing people's privacy, requires therefore to share processors and data in a privacy preserving manner.

In this paper we raise attention to the problem of privacy and security for distributed speech enhancement and propose the new paradigm of privacy preserving distributed beamforming. Using cryptographic techniques, particularly homomorphic encryption, we demonstrate how distributed beamforming techniques can be computed in a privacy preserving manner in the encrypted domain.

Index Terms— wireless sensor networks, distributed algorithms, speech enhancement, encryption

1. INTRODUCTION

Digital speech processing applications play an important role in the world of today. Examples can be found in all sorts of communication devices like mobile telephony, (smart) phone applications, voice controlled devices, hearing aids and cochlear implants. These applications have in common that both speech quality and speech intelligibility degrades under noisy conditions, e.g. at a cocktail party. An effective way to increase quality and intelligibility is to equip these applications with a multi-microphone noise reduction algorithm, e.g., [1–3]. These methods can use spatial and temporal filtering. As a consequence, speech quality and speech intelligibility can be increased [4]. This is in contrast to the single-channel noise reduction algorithms that generally show no speech intelligibility improvement [4, 5] or only very modest improvements, see e.g., [6].

Unfortunately, due to physical limitations, e.g., size and battery power, most applications with multi-microphone noise reduction algorithms are only equipped with a small number of microphones, often at most two or three. This severely limits the performance. The advances in wireless sensor networks (WSNs), allows to develop future applications that can use a much larger number of additional microphones and processors that are present in the environment.

An application scenario would be a room with a pre-installed WSN of microphones, each equipped with a processing unit. In addition, speech processors like hearing aids and mobile phones that are present in the room can connect to the WSN and extend it.

This research is supported by the Dutch Technology Foundation STW.

However, for a large WSN, the conventional centralized multi-channel noise reduction algorithms (for example [1–3]) are neither robust nor scalable, since all processing is done at a single processor and all data in the network needs to be transmitted to this processor. In addition, WSNs may be dynamic as sensor nodes may join or leave the network due to a defect or empty battery, resulting in unpredictable changes in network size and topology.

This requires the development of distributed noise reduction algorithms that work in a decentralized manner. Ideally, they should only make use of local information and be robust in dynamic networks, where microphones may leave or enter the network or even move throughout the environment. This will make them favorable to use in future WSNs over the use of the more conventional centralized multi-microphone noise reduction algorithms that are often constraint to a fixed non-dynamic microphone configuration.

Recently, these requirements led to an increased interest to develop distributed noise reduction algorithms. For example, in [7], a distributed binaural multi-channel Wiener filter (MWF) was presented that uses two hearing aids and converges to the (centralized) binaural MWF presented in [8] for a single target speaker. This work has been generalized in [9] to the distributed adaptive node-specific signal estimation (DANSE) algorithm to handle more sensors and target speakers for distributed estimation of the centralized filter coefficients. In [10] the DANSE algorithm was used in a linearly constrained fashion, to construct a node-specific linearly constrained minimum variance beamformer. Related to this, in [11] a distributed version of the minimum variance distortionless response (MVDR) beamformer was presented exploiting an efficient distributed single constraint generalized sidelobe canceler. Different approaches were followed in [12, 13] and [14], where distributed beamformers were presented based on the randomized gossip algorithm [15] and distributed message-passing algorithms [16, 17], respectively.

While the distributed binaural MWF from [7] involves only two devices each equipped with several sensors and both owned by the same user, the algorithms presented in [9–14] allow the use of many sensors in a distributed fashion not specifically owned by one user. The fact that the WSN can be formed by processors and sensors that are owned by multiple users can lead to severe privacy and security issues. It can lead to the situation that other users of the same WSN know with certainty to which person a hearing aid user intends to listen. Moreover, potentially untrusted people can access the WSN and eavesdrop conversations. More specifically, with distributed beamforming some of the information related to the conversation the user intends to follow will be part of the information flowing through the WSN, e.g., the direction or position of the source of interest by means of the acoustic transfer function, or, intermediate estimates of the signal of interest. Also, as the user might share his own hearing aids or mobile phone as part of the WSN, other users have access to

his processors and microphones as well, while the owner of this device probably does not want to share any details of his conversation or knowledge on the signal he is listening to.

Looking at the recent development on distributed speech enhancement described above and the advances in WSNs, we envision that distributed processing in WSNs will play an important role for future speech communication. At the same time, privacy is one of the most important values of our society. However, despite the increased interest to develop distributed speech enhancement algorithms for WSNs, privacy and security aspects are currently unaddressed. We strongly believe that a key feature for the success and acceptance of WSNs in speech communication systems is the development of algorithms that can guarantee the privacy of the users.

To achieve privacy-protection, the principles of secure signal processing (SSP) [18] [19] can be applied. In SSP, data is encrypted using a homomorphic encryption scheme [20] that allows to perform linear operations on the encrypted data. The decryption key will be generated by one of the users and is unavailable to any other party. This means that no other entity can access the private (encrypted) data. However, using so-called homomorphic properties for linear operations, other entities in the WSN can still process the encrypted data. This allows the WSN to perform noise reduction for speech enhancement, even though the data itself is encrypted and not accessible without intervention of the owner of the decryption key.

In recent years, SSP has proved itself as a promising direction for privacy-enhanced technologies as it is applied for a wide range of signal processing applications, e.g., secure face recognition [21], biometric data matching [22], data clustering [23] and Fourier transformation [24]. In this paper we raise attention to the currently completely untouched, but relevant problem of privacy and security for distributed speech enhancement algorithms and attract awareness to the new and challenging problems that this brings. We demonstrate how existing distributed multi-microphone algorithms can be used within a secure environment using homomorphic encryption.

2. PROBLEM FORMULATION AND NOTATION

To illustrate the problem of privacy preservation for distributed speech enhancement, we use as an example a special case of the distributed delay and sum beamformer (DDSB) presented in [12] in order to estimate a certain target signal from a mix of sources. We consider a situation with multiple users, multiple sources and multiple processing entities, where each processing entity, e.g., a hearing aid, mobile phone or a microphone that is pre-installed in the room, is owned by a (possibly different) user and consists of a microphone and processor. However, notice that the presented framework for privacy preservation is not limited to the method presented in [12].

Consider the situation where a user is interested in a specific source at a certain location. This user considers the remaining sources as noise sources. The DDSB estimates the target signal by processing the microphone data on a frame-by-frame basis in the Fourier domain. Let $Y_i(k,m)$ denote a discrete Fourier transform (DFT) coefficient at entity i, frequency bin k and time-frame m. We assume that all sources are mutually uncorrelated and that the noise sources are additive to the target source, that is,

$$Y_i(k,m) = S_i(k,m) + N_i(k,m), \tag{1}$$

where $S_i(k,m)$ denotes the target (speech) DFT coefficient and $N_i(k,m)$ denotes the noise DFT coefficient at entity i.

The target and noise DFT coefficients are assumed to be independent across time and frequency, which allows us to omit the time and frequency indices for notational convenience. Further, for ease of notation we will use a stacked vector notation, i.e., $\mathbf{Y} = [Y_1, \cdots, Y_M]^T$, with M the number of entities (i.e., wireless sensors in the WSN equipped with a processor) and where the superscript $(\cdot)^T$ denotes transposition of a vector or a matrix. The speech and noise vector $\mathbf{S}$ and $\mathbf{N}$ are defined similarly as $\mathbf{Y}$. Let $\mathbf{d} = [d_1, \cdots, d_M]^T$ be the acoustic transfer function from the speech source to all entities. In order to concentrate on the privacy preserving context, we assume in this work for simplicity a free field situation without damping. This implies that $|d_i| = 1 \; \forall \, i$ and that $d_i \; \forall \, i$ can be computed given the knowledge of the sensor positions and the position of interest. In closed environments, the exact value of d_i also depend on the room acoustics, which we will neglect in this paper for simplicity. Altogether we can then write

$$\mathbf{Y} = \mathbf{S} + \mathbf{N} = S\mathbf{d} + \mathbf{N}, \tag{2}$$

with S the target DFT coefficient at the target location.

Although there are many possible multi-microphone estimators in order to estimate S, we use for illustrative purposes the delay and sum beamformer. Under the above assumptions of a free field without damping, this beamformer is given by

$$\hat{S} = \frac{1}{M} \sum_{i=1}^{M} d_i^* Y_i = \widetilde{Y}_{ave}, \tag{3}$$

with $\hat{S}$ the estimator of S and $(\cdot)^*$ complex conjugation.

2.1. Distributed Delay and Sum Beamformer based on Randomized Gossip

In order to facilitate the discussion on distributed speech enhancement in a privacy preserving context, we briefly summarize in this section the main aspects of the DDSB algorithm presented in [12].

The randomized gossip algorithm [15] can be used to solve consensus problems in a distributed way. Given a connected network of M nodes and initial scalar value $g_i(0)$ at each node i, the randomized gossip algorithm estimates the average value $g_{ave} = \frac{1}{M} \sum_{i=1}^{M} g_i(0)$ using an iterative scheme using only local information and local processing. Let $g_i(t)$ denote the value available at node i in iteration t. Each iteration, a node i and a node j exchange their local information and update their current local estimates as

$$g_i(t) = g_j(t) = (g_i(t-1) + g_j(t-1))/2. \tag{4}$$

Given that the network is connected, this will converge to the average value g_{ave} [15].

The estimator in Eq. (3) can be seen as a (weighted) average. Using the randomized gossip algorithm this average can be computed in a distributed fashion, for each time frame m and frequency bin k. Let $\widetilde{Y}_i(t)$ denote the value available at node i and iteration t. The initial value at node i is then given by

$$\widetilde{Y}_i(0) = d_i^* Y_i. \tag{5}$$

Given this initial value, the randomized gossip algorithm can be used to compute the average $\widetilde{Y}_{ave}$ in Eq. (3) per time frame in an iterative and distributed fashion. This is done by exchanging information across randomly selected node pairs, that is, equivalent to Eq. (4),

$$\widetilde{Y}_i(t) = \widetilde{Y}_j(t) = (\widetilde{Y}_i(t-1) + \widetilde{Y}_j(t-1))/2. \tag{6}$$

The DFT coefficient Y_i, needed to compute the initial value $\widetilde{Y}_i(0)$, can be obtained from the noisy data available at node i. The quantity

d_i^* in $\widetilde{Y}_i(0)$ depends on the (relative location of the) target source and sensor position, and can be provided by the user. In this work, we consider this information about source and sensor positions as private data, as discussed next.

2.2. Targeted Scenario for Privacy Preservation

There are many imaginable scenarios in the distributed speech enhancement context where privacy preservation plays an important role. For illustrative purposes we focus here on the privacy sensitivity of d_i, the acoustic transfer function from source to microphone i. Knowing d_i, anybody with access to the same WSN knows to which source this user is listening to and is able to reconstruct this signal. These d_i's thus convey privacy sensitive information. In our scenario a user wants to keep this information private, while still making use of other (untrusted) entities to estimate a certain desired signal.

Clearly, encryption of $d_i \; \forall \, i$ will prevent information leakage on the specific source the user is interested in. However, encrypting d_i may also make it impossible to perform the necessary mathematical operations in Eq. (3) in distributed manner. In the case of the DDSB, these mathematical operations are scaling of d_i^* and adding several scaled d_i^* values together. This thus leads to two conflicting requirements. While the privacy of the user can be provided by encrypting the d_i's, this encryption may make it infeasible to estimate S in distributed way. Hence, special encryption techniques are needed that perform operations like scaling and addition of the original data, i.e., the plain data, by manipulating the encryptions.

3. HOMOMORPHIC ENCRYPTION

An example of an encryption system that preserves some structure and allows to perform operations like scaling and addition on the plain data by manipulating the encrypted data is the homomorphic Paillier cryptographic system [25]. This is a so-called asymmetric encryption system with two keys, that are, a public key (PK) that can be used for encryption, and a secret key (SK) that can be used for decryption. Let $\mathcal{E}_{pk}$ and $\mathcal{D}_{sk}$ denote the encryption and decryption operations with the public and the secret key, respectively.

The Paillier cryptosystem is *additive*, meaning that multiplication of two Pailler encrypted numbers yields the encryption of the sum of the numbers, that is,

$$\mathcal{D}_{sk}(\mathcal{E}_{pk}(m_1) \cdot \mathcal{E}_{pk}(m_2)) = m_1 + m_2. \tag{7}$$

As a consequence of the additive homomorphism, a number can also be scaled by exponentiating its encryption by a constant, that is,

$$\mathcal{D}_{sk}(\mathcal{E}_{pk}(m)^c) = m \cdot c. \tag{8}$$

Generally, the signal samples involved in speech enhancement problems are within a relatively small range. Homomorphic cryptosystems such as Paillier, on the other hand, have very large message spaces, e.g. 1024 bits, as these cryptosystems rely on mathematical hard problems like nth residues and factorization that requires using very large numbers. An important aspect of the Paillier cryptosystem is the fact that it is *probabilistic*, i.e., in every encryption a random factor is introduced such that encryption of the same number will result in a different encrypted number. The Paillier encryption function for encrypting the number $m \in \mathbb{Z}_n$ is given by,

$$c = \mathcal{E}_{pk}(m) = g^m \cdot r^n \bmod n^2, \tag{9}$$

where $n = p \cdot q$ is a product of two large prime numbers p and q, g is a generator of the group with order n (i.e., $g^n = 1 \bmod n^2$) and can always be chosen as $g = n + 1$, and r is randomly chosen from a specific set of numbers that are co-prime[1] with n. The private key is the tuple (p, q), from which the public key follows as the tuple (n, g). The decryption function for encryption c is defined as

$$\mathcal{D}_{sk}(c) = \frac{L(c^\lambda \bmod n^2)}{L(g^\lambda \bmod n^2)} \bmod n, \tag{10}$$

where $L(u) = (u-1)/n$, $\lambda = \mathrm{lcm}(p-1, q-1)$, with $\mathrm{lcm}(a, b)$ the least common multiple of two integers a and b, i.e., the smallest positive integer that is divisible by both a and b. Notice that as we work with modular arithmetic, a^{-1} is the modular inverse, defined as, $a \cdot a^{-1} = 1 \bmod n$.

As an example, we consider the encryption of the number $m = 5$. In this example, the secret key is chosen as $(p, q) = (3, 5)$ leading to a public key $(n, g) = (15, 16)$. The numbers p and q in this example are relatively small for demonstration purpose. In practice they are of a much higher order for security purposes. Even with this small number $n = 15$, there are already eight numbers r that are co-prime with n, e.g., $r = 7$ or $r = 8$. Encryption of the number 5 can then lead to eight different possible outcomes, e.g.,

$$\mathcal{E}_{pk}(5) = 16^5 7^{15} \bmod 15^2 = 193$$

or

$$\mathcal{E}_{pk}(5) = 16^5 8^{15} \bmod 15^2 = 32.$$

As in practice the number n is very large, it is practically infeasible to determine the plain data m without knowledge of the secret key. To decrypt the encryption 193, we need to compute λ, which is in this example given by $\mathrm{lcm}(2, 4) = 4$. Then,

$$\mathcal{D}_{sk}(193) = \frac{L(193^4 \bmod 15^2)}{L(16^4 \bmod 15^2)} \bmod 15 = \frac{5}{4} \bmod 15 = 5 \; .$$

Notice, that as we work with modular arithmetic, $\frac{1}{4}$ needs to be computed as the modular inverse of 4, that is, $\frac{1}{4} = 4 \bmod 15$.

An important aspect of this encryption scheme is the fact that it operates in a modular domain and that the message m needs to be representable in this domain. This means that m should be an integer number in the range $m \in [0, n-1]$. Notice that negative numbers can still be represented by either using a constant shift, or, recoding the negative number, say $-m$, using its modular inverse.

4. PRIVACY-PRESERVING DISTRIBUTED SPEECH ENHANCEMENT

In this section we will give a description on how a distributed speech enhancement algorithm can be computed in a privacy-preserving manner. As mentioned before, we use as an example a special case of the DDSB presented in [12] that was briefly described in Sec. 2.1.

Speech enhancement algorithms operate on complex, non-integer numbers, while the homomorphic encryption algorithms work on integer numbers from a limited range. This means that the data in a privacy-preserving distributed speech enhancement algorithm based on homomorphic encryption needs to be scaled (depending on the level of significance), quantized, and transformed into real integers. This can be done by scaling and quantizing the data and processing the real and imaginary numbers separately. The scaling and quantization operation will be denoted by $\lceil \cdot \rfloor$.

[1] Two numbers are co-prime if there is no positive divisor that can divide both, except the number one.

As $\widetilde{Y}_i(t)$ in Eq. (6) is complex, it should be evaluated over the real and imaginary parts separately. Let the subscripts $\Re$ and $\Im$ denote the real and imaginary part of a certain variable. Splitting the initial values $\widetilde{Y}_i(0)$ into its real and imaginary parts, we obtain

$$\widetilde{Y}_i(0)_{\Re} = d^*_{i\,\Re} Y_{i\Re} - d^*_{i\,\Im} Y_{i\Im}, \tag{11}$$

and

$$\widetilde{Y}_i(0)_{\Im} = d^*_{i\,\Im} Y_{i\Re} + d^*_{i\,\Re} Y_{i\Im}, \tag{12}$$

respectively.

Applying the DDSB in the encrypted domain requires three steps. At first, the user should send encryptions of d^*_i to every entity i in the network. In the case that reverberation can be neglected, d_i can be obtained based on the positions of the entities relative to the source. Otherwise, it should be estimated, using for example an estimate of the relative transfer function [26]. To focus on the privacy preservation aspect, we put estimation of the quantities d_i in this paper aside, and assume the user has perfect knowledge of them.

Prior to encryption, d^*_i needs to be scaled to integers, and split into its real and imaginary parts leading to $[d^*_{i\,\Re}]$ and $[d^*_{i\,\Im}]$, respectively. Together with the public key, the user sends then the encryptions $\mathcal{E}_{pk}([d^*_{i\,\Re}])$ and $\mathcal{E}_{pk}([d^*_{i\,\Im}])$ to each entity i.

Secondly, in every time frame, each entity has to compute the initial values $\widetilde{Y}_i(0)_{\Re}$ and $\widetilde{Y}_i(0)_{\Im}$ in the encrypted domain. Let $c_{\Re}$ and $c_{\Im}$ denote the quantized values $c_{\Re} = [Y_{i\Re}]$ and $c_{\Im} = [Y_{i\Im}]$, respectively. The initial values $\widetilde{Y}_i(0)_{\Re}$ and $\widetilde{Y}_i(0)_{\Im}$ can then be computed in the encrypted domain as

$$\mathcal{E}_{pk}(\widetilde{Y}_i(0)_{\Re}) = \mathcal{E}_{pk}([d^*_{i\,\Re}])^{c_{\Re}} \mathcal{E}_{pk}([d^*_{i\,\Im}])^{-c_{\Im}}, \tag{13}$$

and

$$\mathcal{E}_{pk}(\widetilde{Y}_i(0)_{\Im}) = \mathcal{E}_{pk}([d^*_{i\,\Im}])^{c_{\Re}} \mathcal{E}_{pk}([d^*_{i\,\Re}])^{c_{\Im}}, \tag{14}$$

respectively.

Thirdly, given the initial values in Eqs. (13) and (14), the randomized gossip algorithm can be run on the encrypted data. To do so, Eq. (6) should be translated to the encrypted domain. Given that nodes i and j communicate in iteration t, they compute

$$\mathcal{E}_{pk}(\widetilde{Y}_i(t)_{\Re}) = \mathcal{E}_{pk}(\widetilde{Y}_j(t)_{\Re}) = \mathcal{E}_{pk}(\widetilde{Y}_i(t-1)_{\Re})^{2^{-1}} \mathcal{E}_{pk}(\widetilde{Y}_j(t-1)_{\Re})^{2^{-1}},$$

and

$$\mathcal{E}_{pk}(\widetilde{Y}_i(t)_{\Im}) = \mathcal{E}_{pk}(\widetilde{Y}_j(t)_{\Im}) = \mathcal{E}_{pk}(\widetilde{Y}_i(t-1)_{\Im})^{2^{-1}} \mathcal{E}_{pk}(\widetilde{Y}_j(t-1)_{\Im})^{2^{-1}},$$

where 2^{-1} denotes the modular inverse of 2. Further, notice that the scaling operation $\lfloor \cdot \rfloor$ should be a multiple of two, related to the number of applied iterations to guarantee that $\widetilde{Y}_i(t)$ stays in the integer domain. After a pre-chosen number of iterations T, the user can decrypt the quantities $\mathcal{E}_{pk}(\widetilde{Y}_i(t)_{\Re})$ and $\mathcal{E}_{pk}(\widetilde{Y}_i(t)_{\Im})$ using the secret key, compensate for the applied scaling, and subsequently compute the estimate $\hat{S}$ and construct the time-domain waveform by computing an inverse DFT followed by an overlap-add.

Although all entities in the network collaborate in estimating $\hat{S}$ for one specific user, no information on the source of interest will leak to these entity, as the secret key is only known by the user.

5. CHALLENGES IN SECURE SIGNAL PROCESSING FOR DISTRIBUTED SPEECH ENHANCEMENT

Homomorphic cryptosystems like Paillier depend on hard problems, which rely on large numbers of hundreds of bits. Consequently, encrypting signal samples, which are usually a couple of bits in size, results in very large numbers, introducing data expansion. Representation of the signal samples in the encrypted domain and their transmission and storage are challenging tasks [27]. Moreover, due to working with large integer numbers in the encrypted domain, operations like exponentiation become significantly expensive in terms of run-time compared to operations in the non-encrypted domain.

Another challenge is that existing practical homomorphic cryptosystems allow to realize either addition or multiplication of non-encrypted numbers, but not both. That is, they allow to either compute $\mathcal{E}_{pk}(a+b)$ or $\mathcal{E}_{pk}(a \cdot b)$. In the current application this means that with respect to Eqs. (11) and (12) we either encrypt the d_i values and consider the noisy DFT coefficients Y_i as constants or vice versa. Multiplying two encrypted numbers is thus problematic in the homomorphic encryption scheme. Realizing such operations in the encrypted domain requires other cryptographic tools, namely secure function evaluation techniques (see e.g. [19]). These techniques mostly consist of interactive protocols. Besides multiplication, this also holds for operations like the logarithm, exponentiation, comparison, division, etc. Unfortunately, existing cryptographic protocols in literature for operations like division and comparison are very generic as they are not designed by taking the signal nature of the data into account. If used naively, the resulting cryptographic protocols for the signal processing algorithms will be impractical in terms of run-time, bandwidth and storage.

A way to reduce the costs is to design tailor-made protocols. This can be achieved in two steps: 1) the signal processing algorithm can be optimized in the numbers of operations, and 2) the cryptographic protocols, particularly the interactive ones with complex operations, can be designed by taking the signal nature of the data into account. To be more precise, a certain signal processing operation does not need to be *exact* and can still be usable in practice. This optimized version of the algorithm with an estimate of the result could be implemented more efficiently in the encrypted domain. For distributed speech enhancement, there are also other challenges like fully distributed microphone systems with no central authority to control the system and extremely limited processing power and storage capacity. This makes it obligatory to design cryptographic protocols using *lightweight* cryptographic tools, which are designed for limited devices such as RFID tags [28]. Nevertheless, to address such problems for speech enhancement, a deep understanding in both cryptography and signal processing is required.

6. CONCLUDING REMARKS

In this work we raised attention to the fact that privacy preservation is a serious issue for distributed multi-microphone noise reduction. We presented a new paradigm of privacy preserving distributed beamforming and proposed a secure distributed speech enhancement technique based on homomorphic encryption. Via an example of a sum and delay beamformer we have shown that privacy preserving beamforming is feasible by processing in the encrypted domain. However, to make privacy preserving beamforming applicable to more advanced distributed multi-microphone noise reduction algorithms, cryptographic protocols should be designed that are adjusted to the specific application in order to reduce communication and computational costs. On the other hand, when designing signal processing algorithms, the potentials and limitations of cryptographic tools have to be accounted for. Thus, to achieve the goal of privacy preserving speech enhancement, an integrated approach between the fields of cryptography and speech signal processing is necessary.

7. REFERENCES

[1] J. Benesty, M. M. Sondhi, and Y. Huang (Eds), *Springer handbook of speech processing*, Springer, 2008.

[2] S. Doclo, *Multi-microphone noise reduction and dereverberation techniques for speech applications*, Ph.D. thesis, Katholieke Universiteit Leuven, 2003.

[3] J. Bitzer and K. U. Simmer, "Superdirective microphone arrays," in *Microphone Arrays: Signal Processing Techniques and Applications*, M. S. Brandstein and D. B. Ward, Eds., Berlin, Heidelberg, New York, 2001, pp. 19–38, Springer-Verlag.

[4] K. Eneman et al., "Evaluation of signal enhancement algorithms for hearing instruments," in *EURASIP Europ. Signal Process. Conf. (EUSIPCO)*, Lausanne, Switzerland, Aug. 2008.

[5] Y. Hu and P. C. Loizou, "A comparative intelligibility study of single-microphone noise reduction algorithms," *J. Acoust. Soc. Amer.*, vol. 122, no. 3, pp. 1777–1786, 2007.

[6] J. Jensen and R. C. Hendriks, "Spectral magnitude minimum mean-square error estimation using binary and continuous gain functions," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 1, Jan. 2012.

[7] S. Doclo, M. Moonen, T. Van den Bogaert, and J. Wouters, "Reduced-bandwidth and distributed MWF-based noise reduction algorithms for binaural hearing aids," *IEEE Trans. Audio, Speech, Language Process.*, vol. 17, no. 1, pp. 38 –51, Jan. 2009.

[8] T. Klasen, T van den Bogaert, M. Moonen, and J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," *IEEE Trans. Signal Processing*, vol. 55, no. 4, pp. 1579 – 1585, April 2007.

[9] A. Bertrand and M. Moonen, "Distributed adaptive node-specific signal estimation in fully connected sensor networks – part I: Sequential node updating," *IEEE Trans. Signal Processing*, vol. 58, no. 10, pp. 5277 –5291, Oct. 2010.

[10] A. Bertrand and M. Moonen, "Distributed node-specific LCMV beamforming in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 233 –246, Jan. 2012.

[11] S. Markovich-Golan, S. Gannot, and I. Cohen, "Distributed GSC beamforming using the relative transfer function," in *EURASIP Europ. Signal Process. Conf. (EUSIPCO)*, Aug. 2012, pp. 1274 – 1278.

[12] Y. Zeng and R. C. Hendriks, "Distributed delay and sum beamformer for speech enhancement in wireless sensor networks via randomized gossip," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2012.

[13] Y. Zeng and R. C. Hendriks, "Distributed delay and sum beamformer in regular networks based on synchronous randomized gossip," in *Int. Workshop Acoustic Signal Enhancement (IWAENC)*, sept. 2012.

[14] R. Heusdens, G. Zhang, R. C. Hendriks, Y. Zeng, and W.B. Kleijn, "Distributed MVDR beamforming for (wireless) microphone networks using message passing," in *Int. Workshop Acoustic Signal Enhancement (IWAENC)*, 2012.

[15] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. on Information Theory*, vol. 52, no. 6, pp. 2508 – 2530, June 2006.

[16] G. Zhang and R. Heusdens, "Convergence of generalized linear coordinate-descent message-passing for quadratic optimization," in *IEEE Int. Symposium on Information Theory Proceedings (ISIT)*, 2012, pp. 1997–2001.

[17] G. Zhang and R. Heusdens, "Linear coordinate-descent message-passing for quadratic optimization," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2012, pp. 2005–2008.

[18] Z. Erkin, A. Piva, S. Katzenbeisser, R. L. Lagendijk, J. Shokrollahi, G. Neven, and M. Barni, "Protection and retrieval of encrypted multimedia content: When cryptography meets signal processing," *Eurasip Journal on Information Security*, vol. 2007, 2007.

[19] R. L. Lagendijk, Z. Erkin, and M. Barni, "Encrypted signal processing for privacy protection," *IEEE Signal Process. Mag.*, Jan. 2013, in press.

[20] C. Fontaine and F. Galand, "A survey of homomorphic encryption for nonspecialists," *EURASIP Journal on Information Security*, vol. 2007, 2007.

[21] A. Sadeghi, T. Schneider, and I. Wehrenberg, "Efficient privacy-preserving face recognition," in *12th International Conference on Information Security and Cryptology (ICISC)*. Dec. 2009, LNCS, Springer.

[22] M. Blanton and P. Gasti, "Secure and efficient protocols for iris and fingerprint identification," in *European Symposium on Research in Computer Security*, Leuven, Belgium, Sept. 2011.

[23] Z. Erkin, T. Veugen, T. Toft, and R. L. Lagendijk, "Privacy-preserving user clustering in a social network," in *1st IEEE Workshop on Information Forensics and Security (WIFS)*, 2009, pp. 96–100.

[24] T. Bianchi, A. Piva, and M. Barni, "On the implementation of the Discrete Fourier Transform in the encrypted domain," *IEEE Trans. Inf. Forensics Security*, vol. 4, no. 1, pp. 86–97, Mar. 2009.

[25] P. Paillier, "Public-Key Cryptosystems Based on Composite Degree Residuosity Classes," in *Advances in Cryptology — EUROCRYPT '99*, J. Stern, Ed. May 1999, vol. 1592 of *LNCS*, pp. 223–238, Springer.

[26] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. on Signal Processing*, vol. 49, no. 8, pp. 1614–1626, 2001.

[27] T. Bianchi, A. Piva, and M. Barni, "Composite signal representation for fast and storage-efficient processing of encrypted signals," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 1, pp. 180–187, Mar. 2010.

[28] D. Maimut and K. Ouafi, "Lightweight cryptography for RFID tags," *IEEE Security Privacy*, vol. 10, pp. 76–79, 2012.