Implementing and developing secure low-cost long-range system using speech signal processing

Samer Alabed[1], Amer Alsaraira[1], Nour Mostafa[2], Mohammad Al-Rabayah[2], Yehia Kotb[2], Omar A. Saraereh[3]

[1]Department of Biomedical Engineering, School of Applied Medical Sciences, German Jordanian University, Amman, Jordan
[2]College of Engineering and Technology, American University of the Middle East, Kuwait
[3]Department of Electrical Engineering, Faculty of Engineering, The Hashemite University, Zarqa, Jordan

Article Info	ABSTRACT
	In the proposed work, we present a secure low-cost speech communication system for long-distance communication. The system utilizes long range (LoRa) technology to transmit speech signals. LoRa technology uses spread-spectrum modulation to enable long-range communication with low power consumption. LoRa modulation allows for data transfer at a slow speed, typically below 22 kbps, which makes it infeasible for transmitting speech. To address this limitation, we suggest a speech coding technique that reduces the overall data rate of speech signals to below 7.5 kbps. This lower rate is more compatible with the LoRa module and ideal for transmitting speech. Moreover, this technique can improve the LoRa transmission range. Additionally, we have developed an encryption-decryption method to ensure the privacy of the messages and prevent unauthorized access by third parties.

Corresponding Author:

Samer Alabed
Department of Biomedical Engineering, School of Applied Medical Sciences
German Jordanian University
Amman, 11180 Jordan
Email: samer.alabed@gju.edu.jo

1. INTRODUCTION

Low-cost long-range wireless communication systems are becoming increasingly popular due to their ability to transmit information over long distances without the need for physical cables or wires [1]-[5]. These systems are used in a wide range of applications, including military and defence, remote sensing, smart city applications, healthcare and medical applications, environmental monitoring, agriculture and telecommunications. There are several types of long-range wireless communication systems, including satellite communication systems, radio frequency communication systems, and microwave communication systems. These systems use different types of technologies and operate at different frequencies [6]-[8]. Long range (LoRa) is a wireless communication technology that provides a long-range, low-power solution for the internet of things (IoT) devices [9]-[11]. LoRa uses spread-spectrum modulation, which allows for long-range communication with low power consumption. It operates in the unlicensed radio frequency bands, which means that it is available for use without any licensing fees. LoRa devices can communicate over distances of several kilometres, making it an ideal solution for IoT devices that need to be deployed over a large area [11]-[13]. LoRa technology offers low power consumption, which is essential for battery-powered IoT devices. It also provides high network capacity and can support a large number of devices, making it suitable for large-scale IoT deployments. LoRa is a proprietary low-power, wide-area network (LPWAN) technology developed by Semtech Corporation [9]-[12]. Currently, LoRa technology is utilized in various domains, including but not

limited to the tracking of assets, farming, industrial automation, smart homes, healthcare systems, water management, environmental monitoring, and smart city applications, which comprise smart lighting, parking management, waste management, and air quality monitoring.

On the other hand, a secure communication system is one that ensures the privacy and confidentiality of the transmitted data [14]-[15]. The system should be designed in such a way that unauthorized parties cannot access the data, and the data should be transmitted without alteration. An example of low cost and secure communication system is the use of end-to-end encryption. This method encrypts data at the source device and decrypts it at the destination device, making it difficult for third parties to intercept the data. Many messaging apps, such as WhatsApp and signal, use end-to-end encryption, making them secure and affordable communication tools. These systems ensure the privacy and confidentiality of transmitted data while remaining affordable and accessible to a wide range of people [16]-[19].

As it is mentioned, LoRa uses spread-spectrum modulation to enable long-range communication with low power consumption. This modulation scheme allows LoRa devices to transmit data at a low bit rate while using very little power [9]-[12]. The low bit rate of LoRa is typically in the range of a few kilobits per second (kbps). This means that it is not possible to send a speech signal over LoRa modules since its data rate is around 64 kbps). Currently, LoRa technology is not being used for transmitting speech signals due to the fact that the data rate required for speech signals is higher than what LoRa modules can handle. As a result, the objective of this research paper is to create a speech coding method that can decrease the data rate of speech signals to a few kilobits per second so that they can be transmitted through LoRa modules. Additionally, the paper aims to develop an encryption-decryption method to secure the message from unauthorized access. There are several encryption methods available in literature that could be used to achieve this goal; however, this paper concentrates on developing a fast and reliable encryption-decryption method that uses LoRa modules to ensure message privacy and prevent unauthorized access. Finally, this work aims to be budget-friendly, so it can be accessible to the public, and to find the most efficient and effective solutions through various experiments.

Chirp spread spectrum (CSS) is a spread-spectrum technique used to encode information using linear frequency-modulated chirp pulses which are sinusoidal signals that changes its frequency over time [19]-[20]. CSS utilizes the entire bandwidth to transmit signals, providing maximum sound transmission. It is an ideal method for long distance communication due to its low power consumption. This technique is used in pulse compression radar. The CSS system offers several advantages, including excellent range-to-data rate scalability, low power consumption, and low latency, which results in maximum efficiency. The CSS's scalability enables easy adjustment of data transfer over a specific distance, making it a versatile option for communication.

To make digital speech signals more efficient for storage and transmission, speech coding techniques are utilized. There are two primary types of techniques: waveform-based and model-based. Waveform-based approaches directly analyze the speech signal and aim to reduce redundant information by compressing the waveform through methods such as lowering the sampling rate or bit depth, or using techniques like pulse-code modulation or delta modulation. Model-based approaches, on the other hand, study the speech characteristics and employ mathematical models to represent the speech signal, which can be achieved through techniques like linear predictive coding (LPC) [21]-[29].

The LPC technique is commonly used in low-bit-rate speech coding applications, such as voice-over-internet protocol (VoIP) and mobile communications, and works by using a set of linear predictive coefficients to model the speech signal. These coefficients represent the spectral envelope of the speech signal and can be employed to recreate the speech signal at the receiver. Ultimately, speech coding techniques play a crucial role in optimizing the transmission and storage of speech signals across a wide range of applications, including telecommunications, multimedia, and voice recognition systems [21]. From Figure 1, the traditional LPC method starts by dividing the speech signal into several frames in the range of 30-100 frames per second. These frames are later processed using different filtering stages to produce the speech output. This process includes LPC analysis and pitch detection. The segmented signals are then evaluated against certain parameters to ensure they produce an accurate voice signal before being combined and sent. We are focusing in this article to propose an LPC technique to be used in the LoRa technology to compress speech signals before transmission to conserve power and increase the range of communication.

2. RESEARCH METHOD

The LPC model as shown in Figure 2, divides the speech signal into two parts: the analysis component and the synthesis component. In the analysis part, as shown in Figure 2(a), the residual signal is generated by using the speech signal as an input. On the other hand, in the synthesis component Figure 2(b), the decoder takes the residual signal as input to reproduce the speech signal. Figure 2(a) produces the residual signal e(n), which is also referred to as the error or excitation signal, and is crucial in ensuring that the synthesized signal

accurately corresponds to the original signal. If the error signal is not applied as an input during the synthesis stage or if a synthesis other than 1/A(z) where A(z) represents the transfer function of the analysis filter is used, the reproduced speech signal produced by Figure 2(b) will not match the original signal. It should be noted that s(n) is used to denote the original speech signal, and s'(n) is used to represent the synthesized signal.
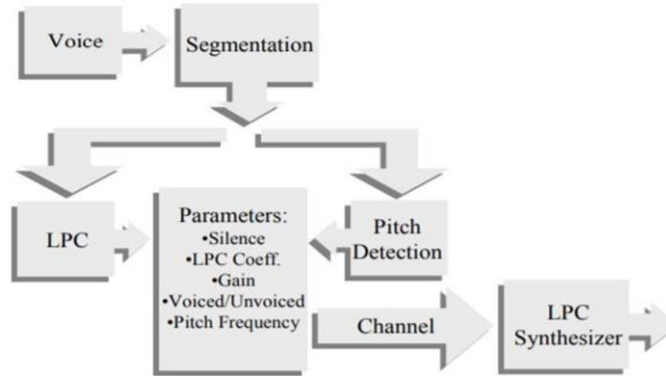

Figure 1. Flowchart of the traditional linear predictive coding process
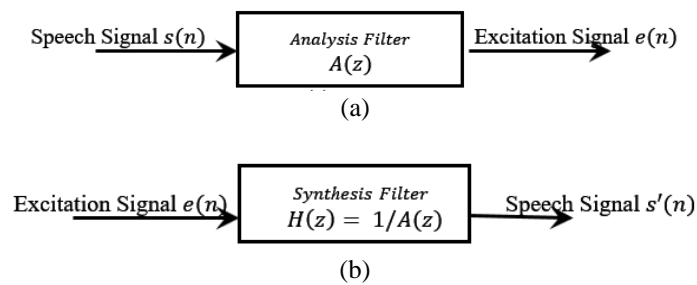

Figure 2. Linear predictive coding model (a) coder and (b) decoder

The production of human speech involves two crucial components: The excitation source and the vocal tract configuration. An all-pole transfer function $H(z) = 1/A(z)$ is used to model the vocal tract, which is activated by a glottal excitation (Residual) signal $e(n)$ in the form of a discrete-time sequence to create the speech signal $s(n)$. Speech signals are not constant and can only be considered stable for brief periods of time (20-40 ms) [25], [26]. To model speech production accurately, it is essential to consider both the excitation source and the vocal tract. In linear predictive coding synthesis, the optimal excitation is the residual signal which is the difference between the predicted and actual signals.

$$e(n) = s(n) - s'(n) = s(n) + \sum_{k=1}^{p} \alpha_k s(n-k) \tag{1}$$

In the traditional LPC system, the excitation signal is represented either as a series of pitch-related impulses for voiced speech frames or as a random noise sequence for unvoiced speech frames. However, alternative techniques like, multi-pulse linear prediction, residual excited linear prediction and regular-pulse excitation [25] take different approaches for modelling the excitation signal and deviating from traditional linear predictive coding. In the proposed innovative system, we aim to enhance the traditional LPC method by lowering the data rate of the original voice while retaining its perceived quality. Our proposed approach improves upon classical linear predictive coding by utilizing a model of the tenth order for the voiced frames and a model of the eighth order for the unvoiced frames, and to determine the needed parameters for linear prediction of each frame by sampling the speech signal at 8 kHz. The ultimate goal is to maintain the speech signal perceptual quality while reducing its bit rate. The original speech signal is broken down into individual frames, and each frame is processed to obtain the residual signal. This residual signal is then divided into 6 smaller sub-frames. An energy classification is performed on the sub-frames to determine the appropriate number of pulses for each sub-frame, with sub-frames that have higher energy receiving a larger number of

pulses. Moreover, the main frames are categorized as either voiced or unvoiced, with unvoiced frames having fewer pulses compared to voiced frames. This sub-framing and categorization process optimizes the selection of pulses, leading to a low data rate and a recovered speech signal with a good quality.

The encoder's analysis system employs linear prediction to calculate the filter parameters α_k and the excitation. The use of linear prediction parameters to represent the vocal tract is considered adequate and results in high-quality speech for voiced and unvoiced speech frames. A (10th order) linear predictive coding model is employed for voiced frames, while another (8th order) model is utilized for unvoiced frames since a lower order model suffices for describing the unvoiced speech spectrum. The calculation of the filter parameters and the excitation is typically performed every 20-40 ms (160-320 samples) for speech sampled at 8 kHz. The excitation signal is considered as the optimal input signal at the receiver side and many linear predictive coders rely on efficiently coding this signal. The excitation signal contains certain components or characteristics that are not captured by the analysis part, including phase, pitch, and zeros produced by noise parts. In the proposed model, the residual signal is encoded by selecting the most important pulses instead of encoding all the pulses. This is done by splitting the frame into sub-frames, classifying them based on energy, and choosing the number of pulses accordingly.

The new encoder model is illustrated in Figure 3. The original speech signal in the time domain is first divided into main frames. These main frames are then classified as either voiced or unvoiced based on an energy threshold, with voiced frames having higher energy than the threshold and unvoiced frames having lower energy. The residual signal will be produced by filtering the main voiced frame, which is then divided into 6 sub-frames. The sub-frames are classified based on their energy content, as the higher energy frames will be given more pulses than lower energy frames. If the main frame is unvoiced, the same process is followed, but without the energy classification, and all sub-frames will be given the same number of pulses equal to that given to the sub-frame with the lowest energy in the voiced frame. By dividing the residual signal into 6 sub-frames and performing the energy and voiced/unvoiced classifications, the best pulses are selected to achieve the needed high quality and low bit rate of the reconstructed speech signal.
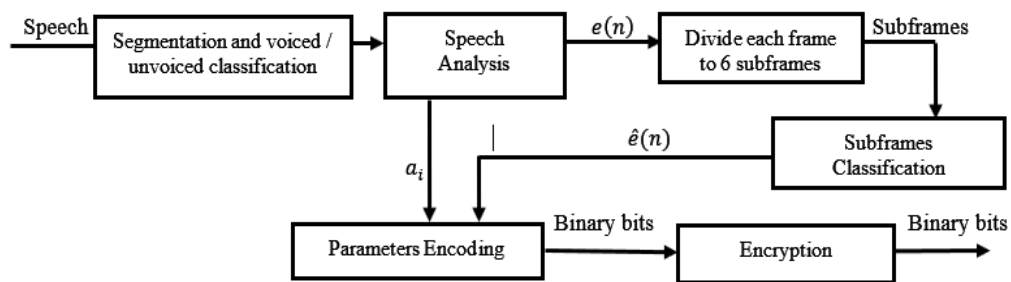

Figure 3. Encoder

At the decoder side, both the coefficients and the excitation signal obtained from the analysis system will be used to excite the filter of the synthesis system to obtain the reconstructed speech signal, as illustrated in Figure 4. The process starts with determining whether the frame is voiced or unvoiced. To regenerate the excitation signal, we use a significant number of samples, which is used to excite the 10th order LPC-based filter when the frame is voiced. For unvoiced speech frame, we use a few samples to regenerate the excitation signal. To ensure energy matching between reconstructed speech frames and original frames, the reconstructed frames are multiplied by a gain. Subsequent frames are multiplied by a Hanning window, followed by an overlap and add operation to combine the frames and produce the entire speech signal.
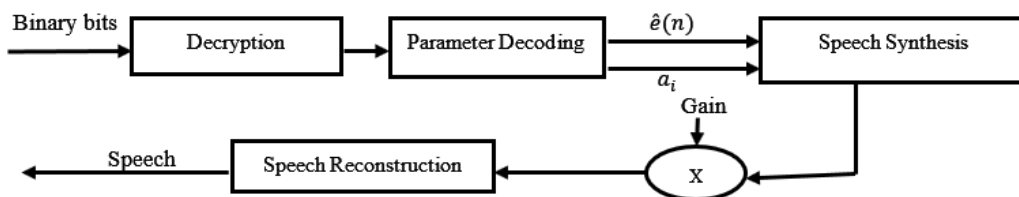

Figure 4. Decoder

The proposed system aims to present a powerful coding system with a bit rate ranging between 5 and 7.5 kbps depending on the energy threshold. The quality of the recovered speech signal is considered to be good at bit rates higher than 5 kbps, as the primary focus is on coding the crucial residual components that preserve the perceptual quality. The next part of this section will discuss the quantization procedure for the pulse amplitudes and line spectrum pairs (LSPs).

Note that the original speech signal is normalized prior to the speech coding process. Thus, the values of the residual signal vector $e = [e_1, e_2, \ldots e_N]$ are very small where its maximum value, $e_{max} \geq e_i \forall i \in \{1, 2, 3, \ldots N\}$, is much less than one, i.e., $e_{max} \ll 1$. The aim of encoding the residual signal's amplitudes is to minimize the quantization error and reconstruct the speech signal of high quality. The suggested approach to encode the residual signal's samples involves determining the suitable scale such that the amplitude's lengths lie within the interval (1-16) as demonstrated in Table 1. This enables us to utilize 4 bits (including the sign bit) for representing the amplitudes after scaling. The procedure can be outlined as: Given that, N is the number of samples to be considered, $e = [e_1, e_2, \ldots e_N]$, then the proper scale will be determined such that the magnitude is between 1 and 16, as illustrated in Table 1. After scaling, the magnitudes can be represented by four bits (including the bit of the sign), which allows utilizing the full resolution. This method can be explained in the below steps:

a.   Find $g_{max} = \log_{16}(e_{max})$

b.   If $g_{max} > -2$, then the first 2 bits in the frame are [00].

c.   Else if $-3 < g_{max} < -2$ then the first 2 bits in the frame are [01].

d.   Else if $-4 < g_{max} < -3$, then the first 2 bits in the frame are [10].

e.   Else, use 11 as the first 2 bits in the frame.

f.   If the first and second bit are 00, multiply the magnitudes of that segment (e) by 16^2. If the first and second bits are 01, multiply the magnitudes of that segment (e) by 16^3. If the first and second bits are 10, multiply the magnitudes of that segment (e) by 16^4. If the first and second bits are 11, multiply the magnitudes of that segment (e) by 16^5, as illustrated in Table 1.

g.   Subtract one from the magnitudes in that segment, then round the results to be in the range (0 to 15) and finally convert the results into 4-bit binary numbers.

In order to minimize the quantization error and preserve reconstruction of speech signal with high quality, we also require a method for encoding the LPC coefficients. To achieve this, we utilize the line spectral frequencies (LSFs) and LSPs, by defining a clearly dynamic range, instead of the LPC coefficients in the time domain. The LSF or LSP coefficients are ideal for robust and efficient quantization of LPC parameters. It is worth noting that the low-frequency parts are represented by the first LSFs, while the high-frequency parts are represented by the last LSFs. Because components with high frequencies have less influence on the perception of speech, fewer bits can be assigned to the higher LSFs compared to the lower LSFs, which correspond to lower frequency parts. This leads to a reduced data rate with largely preserving the quality of the speech signal.

Table 1. Scale code

Scale code	Range prior adjusting (Linear)	Range prior adjusting ($\log_{16}$)	Scale number	Range after adjusting
Rare	$\infty - 16^{-1}$			
00	16^{-1} - 16^{-2}	[-1 , -2]	16^2	1-16
01	16^{-2} - 16^{-3}	[-2 , -3]	16^3	1-16
10	16^{-3} - 16^{-4}	[-3 , -4]	16^4	1-16
11	16^{-4} - 16^{-5}	[-4 , -5]	16^5	1-16
Rare	$0 - 16^{-5}$			

The LSP coefficients in the frequency domain indicate the poles of the analysis filter H(z) or the zeros of the synthesis filter A(z), making them suitable for accurate and effective quantization of the LPC coefficients. Furthermore, the LSPs are linked to the line spectral frequencies and are more concentrated around formants. The LSPs also have localized spectral sensitivity, so encoding the LPC parameters as LSPs enables efficient and robust quantization. The LSPs are a sorted group of numbers ranging from 0 to π and arranged in an ascending sequence. A method has been developed for encoding the LSP parameters in an efficient manner with regards to quantization. The method is outlined in the following order:

a.   Divide the LSP parameters by π, to produce LSP parameters from zero to one.

b.   Modify the scale of the first LSP parameter to be converted using 5-bit binary number.

c.   Transform the output to a binary number of 5 bits.

d.   Subtract the result found in step 2 from all other LSP parameters of that segment.

e.   Round the output obtained in the former step.

f.   Transform the output obtained in the former step to a binary sequence.

g. Subtract the value found in step 5 from all other LSPs in the same frame.
h. Repeat steps five to seven til all LSP parameters have been processed. The number of bits employed for LSP parameters two to ten is {4/4/4/4/3/3/3/3/2} for voiced frames and {4/3/3/3/2/2/2} for unvoiced frames.

The speech signal is broken down into main frames, each lasting 36 milliseconds (288 samples). The residual signal will be next found by filtering those main frames and then further divided the voiced frames into 6 smaller sub-frames, with 6 milliseconds duration (48 samples). This filtering process will be applied using either ten LPC parameters for voiced frames or eight LPC parameters for unvoiced frames. For voiced frames, the energy levels of the sub-frames are sorted and the two sub-frames with the highest energy levels are given 12 pulses each, resulting in a total of 24 pulses or 1/4 of the 48 pulses (pulse rate of 1:4).

The method for choosing the 12 pulses is as: The residual signal in each of the two sub-frames with the highest energies is divided into 4 residual signals. The first residual signal includes pulses 1, 5, 9, and 13, the second residual signal includes pulses 2, 6, 10, 14, etc., the third residual signal includes pulses 3, 7, 11, and 15, and the fourth excitation signal includes samples 4, 8, 12, and 16. Then, the excitation signal with the largest energy is selected. If the largest energy is in the first residual (excitation) signal, the lag and lead bits are 00. If it is in the second residual signal, the bits will be 01. For the third residual signal, the binary bits are 10. Finally, if it is in the fourth excitation signal, the binary bits are 11. The next two sub-frames with the greatest energy will receive eight samples, which is around 1/6 of the 48 samples (sample rate 1:6). The selection process is the same as for the 12 pulses. However, this time the residual signal in the sub-frame is separated into 6 residual signals. Thus, 3 bits are necessary for the lead and lag information. Lastly, the last two sub-frames with the least energy will each be assigned 6 pulses, which is equivalent to 1/8 of the 48 pulses (pulse rate 1:8). The procedure for selecting these pulses is the same as for the previous 12 and 8 pulses.

The process for selecting the pulses in a voiced main frame involves assigning 12 pulses to the two highest energy sub-frames, 8 pulses to the next two highest energy sub-frames, and 6 pulses to the two lowest energy sub-frames. The lag and lead information for the 12 pulses selection requires 2 bits each. The 8 pulses selection requires 3 bits each, while the 6 pulses selection requires 4 bits each. This results in a total of 18 bits for the lag and lead information in a voiced frame. Each pulse is quantized using 4 bits and the scale code uses 2 bits. The LPC coefficients, represented as LSPs, require 35 bits for voiced frames with 10 coefficients and 24 bits for unvoiced frames with 8 coefficients. Finally, 5 bits are allocated for the energy value which represents the original frame energy and is used in the decoding process.

The number of bits per voiced frame is 269 and per unvoiced frame is 180. In case of choosing a threshold value to have 50% of the frames voiced, the data rate (R) is calculated as: R = (269 + 180)/2 * 27.77 = 5,750 bits per second. In the worst-case scenario of choosing very large threshold value, all frames will be voiced, the total bit rate is R = 263 * 27.77 = 7305 bits per second. Thus, the system provides a high-quality speech reconstruction with a data rate that is typically lower than 7.5 kbps. The data rate capability of the LoRa module is even higher, with a capacity of up to 22 kbps, exceeding the requirement of the worst-case scenario.

3.   RESULTS AND DISCUSSION

The design of the system is as follows: An input in the form of speech is taken in MATLAB. The MATLAB code then processes the input, compresses it into a suitable size, and encrypt the binary bits for transmission. The compressed encrypted data is sent to the Arduino serial port and divided into specific-sized packets. The Arduino communicates with the LoRa module to start transmitting the encrypted data. The LoRa module then spreads the signal using CSS modulation. On the receiving end, another LoRa module scans for incoming packets and receives them when detected. Upon detection of a packet header, the receiver sends an acknowledgment message, allowing the transmitter to send the entire packet. The process results in a good-quality reconstructed signal with a data rate of less than 7.5 kbps. The LoRa module has a data rate capability of up to 22 kbps, which is more than enough to handle the data rate required for the module. Lower data rates, below 22 kbps, correspond to an increase in transmission range. The LoRa technology offers various spreading factors (SF) and bandwidths to optimize the balance between data rate and range. Using higher SF leads to longer range but lower data rate, whereas wider bandwidth results in shorter range but higher data rate. The article aims to decrease the data rate of speech signal significantly, below 22 kbps, to enhance transmission range. Higher SF LoRa technology can enhance the range of transmission for speech signals, which is advantageous for scenarios that require long-range communication, such as in emergency response situations or public safety. An example is emergency responders who can use high SF LoRa to transmit voice communications over long distances, allowing them to coordinate their actions effectively in times of disaster. Furthermore, high SF LoRa is suitable for remote communication in locations with poor cellular coverage and for communicating with equipment in remote areas. Ultimately, using higher SF LoRa to transmit speech signals can optimize long-range communication in different fields.

3.1. Circuit dagrams

From Figure 5, the Arduino board is connected to MATLAB, while the microcontroller is linked to the LoRa module. Similar setup is utilized for the transmitter component, as shown in Figure 5, with the transmitter circuit linked to MATLAB and the LoRa module linked to the Arduino board. The code for the transmitter is comparable to that of the receiver, but with the primary distinction being that the transmitter is focused on sending the data, whereas the receiver is designed to get the data and play it through speaker output.
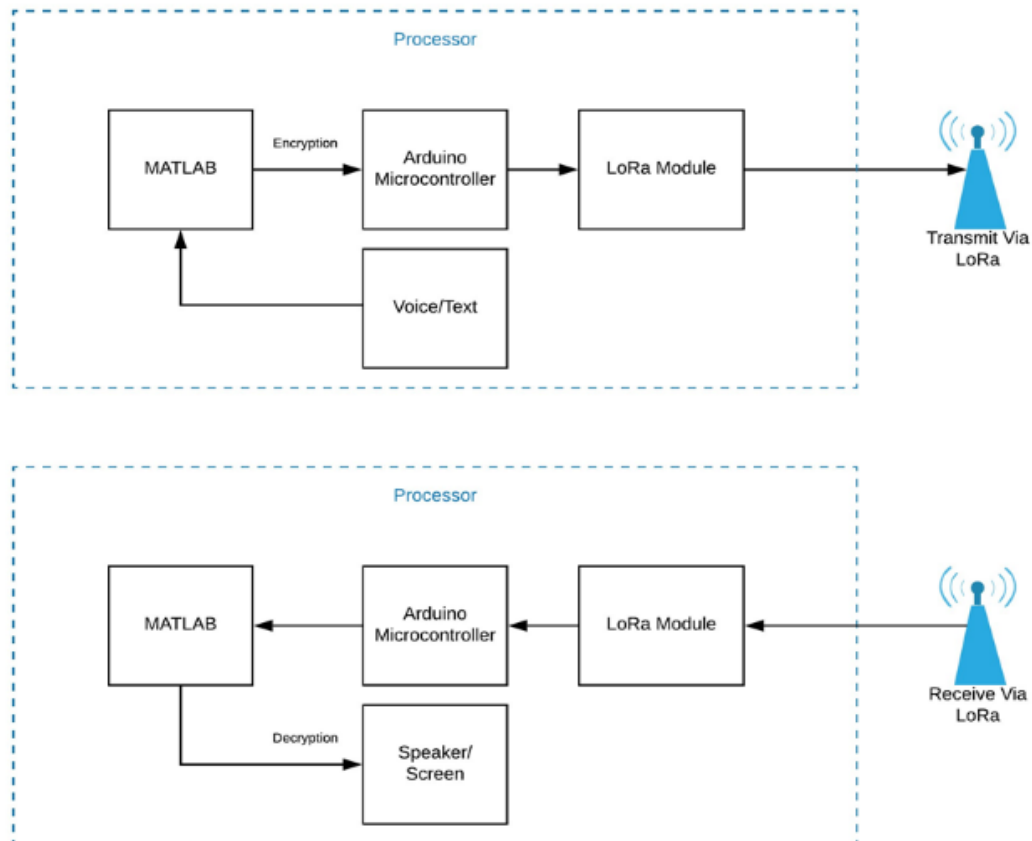
Figure 5. High-level design

3.2. Printed circuit board (PCB)

The next step after a successful breadboard design was to create a printed circuit board (PCB) for the single LoRa module. If multiple modules were to be used, a different design would be necessary for the multiple-input multiple-output system. Figure 6 shows the design of the PCB for one module which features all connected grounds and individual pads for each pin. Each pad has its own copper trace with a hole at the end for easy connection to the microcontroller using male or female pins. The 3.3 V Vcc pin is also indicated. Figure 7 and Figure 8 show the receiver and transmitter schematics for the MATLAB-Arduino interaction.
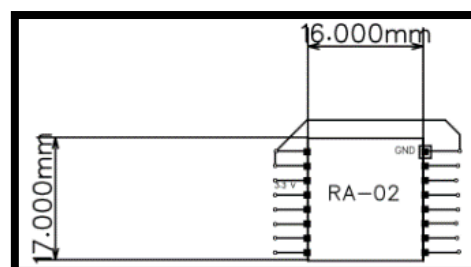
Figure 6. Single LoRa PCB module

3.3. Encryption technique

The proposed speech coding technique discussed in section 2 resulted in a reduced data rate, which made it easier to transmit the data. The generated output bits were divided into packets of 128 bits each, as shown in the message in:

$$m = 11100011\ 10100011\ 10101100\ 11110000\ 00001111\ 01110011\ 00011100\ 11100011$$
$$10101010\ 01010101\ 00101110\ 11010000\ 00110011\ 10101100\ 00111100\ 11000011 \qquad (2)$$

then, we use a certain key which contains 128 bits as shown in the key below:

$$k = 01010011\ 11001100\ 01011100\ 10100011\ 01101111\ 01010011\ 01011100\ 10101011$$
$$10011010\ 00110101\ 01001110\ 10110000\ 10101011\ 00101011\ 00101110\ 11010010 \qquad (3)$$

then, XORing both of the message (m) given in (2) and the key (k) given in (3), we will get the encrypted message:

$$y = m \oplus k = 10110000\ 01101111\ 11110000\ 01010011\ 01100000\ 00100000\ 01000000$$
$$01001000\ 00110000\ 01100000\ 01100000\ 01100000\ 10011000\ 10000111$$
$$00010010\ 00010001 \qquad (4)$$

the receiver will use the key given in (3) to decrypt the received encrypted message given in (4) by applying XOR on it, thus obtaining the original message. Therefore, the original message is given by:

$$m = y \oplus k = 11100011\ 10100011\ 10101100\ 11110000\ 00001111\ 01110011\ 00011100$$
$$11100011\ 10101010\ 01010101\ 00101110\ 11010000\ 00110011\ 10101100$$
$$00111100\ 11000011 \qquad (5)$$

where m denotes the received encrypted message, k denotes the key and $\oplus$ stands for the XOR operation. Note that the received encrypted message is the one received from the transmitter and the key is the same used by the transmitter to encrypt the original message and the XOR operation will restore the original message. Figure 9 shows two speech signals, one in red and the other in blue. The red signal is the reconstructed version using LPC technique, while the blue signal is the original. Based on the figure, it appears that the two signals are quite similar or closely resemble each other.

3.4. Realization

LoRa transceiver was not commonly used for transmitting speech signals, but instead was primarily utilized for sending sensor readings over radio frequency depending on the specific work's needs. Hence, we proposed the implementation of a speech coding technique to reduce the data rate and simplify the transmission process. To summarize, the system had a high-level design which started with the input speech taken in MATLAB which was then encrypted, compressed, and transmitted to the Arduino via the serial port.

The transmitted data was divided into packets and sent through the LoRa module using CSS modulation. The receiver, LoRa module, scanned for packets, received them and sent an acknowledgment message. The received data was then processed by the Arduino and displayed on a speaker output. The design was optimized by creating a printed circuit board design for the LoRa module.

Overall, the prototype demonstrated the potential for using LoRa technology for speech signal transmission. According to section 2, the process of transmitting speech signals involves multiple steps. The objective of the work is to develop a speech synthesizer that could convert speech signals into binary bits, encrypt them, and transmit the result. However, instead of making hardware modifications, the work relied on using MATLAB for recording and processing speech signals.

The prototype started with estimating the expected number of bits that would be transmitted and received. The recorded speech suffers from high data rate, so it was compressed to reduce its data rate. The compression process involved various steps to make it suitable for transmission and reception as discussed in section 2.

The ideas proposed in the paper are mainly applicable in the context of the LoRa Alliance, LPWAN, IoT, and machine-to-machine (M2M) communications. The main goal of the research paper is to create a speech coding technique and an encryption-decryption method that can be utilized to send speech signals through LoRa modules, which are commonly employed in IoT and M2M applications. The proposed methods would enhance the reliability, privacy, and accessibility of communication in LoRa-based IoT and M2M systems by decreasing the data rate of speech signals and safeguarding them using an effective encryption

method. Additionally, the proposed concepts strive to be cost-effective and efficient through a variety of experiments, making it easier to apply LoRa-based systems in different areas such as smart cities, agriculture, and healthcare systems.
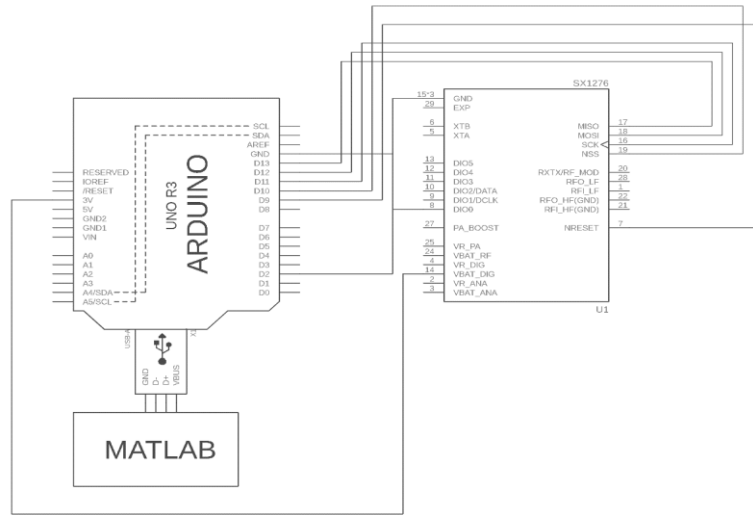

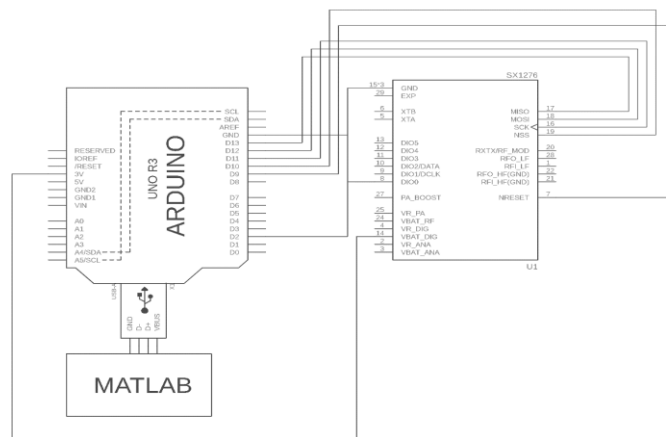Figure 7. Receiver schematic for the MATLAB-Arduino interaction

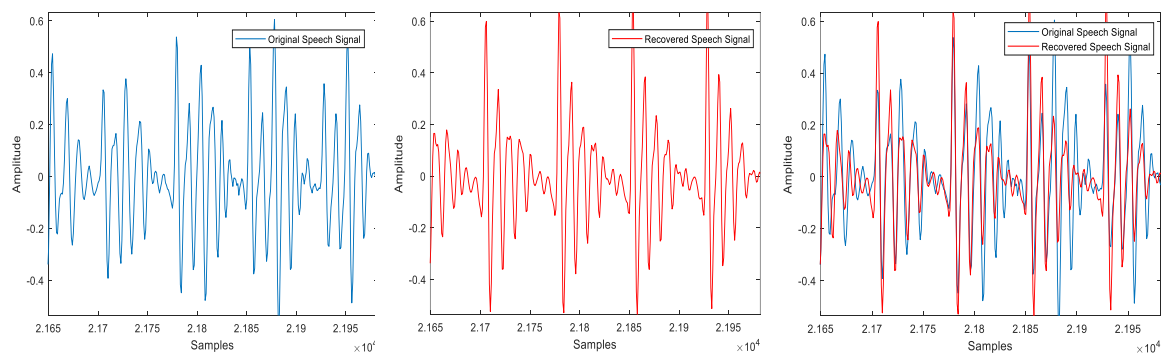Figure 8. Transmitter schematic design for the MATLAB-Arduino interaction

Figure 9. Graphs depicting the original signal, recovered signal, and a comparison of both signals

4.   CONCLUSION

Our proposed work introduces a speech communication system that is both secure and low-cost for long-distance communication. To transmit speech signals, the system makes use of LoRa technology, which utilizes spread-spectrum modulation to enable long-range communication while consuming minimal power. Despite the low data rate of LoRa devices, which typically range in the kilobits per second, our system employs a speech coding technique to reduce the data rate of speech signals to a level that is compatible with LoRa modules. As a result, speech signals can be transmitted through LoRa devices. Furthermore, we have developed an encryption-decryption method to ensure message privacy and prevent unauthorized access from third parties.

REFERENCES

[1]   G. E. John, "A low cost wireless sensor network for precision agriculture," in *Proceedings - 2016 6th International Symposium on Embedded Computing and System Design, ISED 2016*, Dec. 2017, pp. 24–27, doi: 10.1109/ISED.2016.7977048.

[2]   F. E. M. Covenas, R. Palomares, M. A. Milla, J. Verastegui, and J. Cornejo, "Design and development of a low-cost wireless network using IoT technologies for a Mudslides monitoring system," in *2021 Ieee Urucon, Urucon 2021*, Nov. 2021, pp. 172–176, doi: 10.1109/URUCON53396.2021.9647379.

[3]   M. Ballerini, T. Polonelli, D. Brunelli, M. Magno, and L. Benini, "Experimental evaluation on NB-IoT and LoRaWAN for industrial and IoT applications," in *IEEE International Conference on Industrial Informatics (INDIN)*, Jul. 2019, vol. 2019-July, pp. 1729–1732, doi: 10.1109/INDIN41052.2019.8972066.

[4]   M. E. Irfan Bin Edi, N. Emileen Abd Rashid, N. N. Ismail, and K. Cengiz, "Low-Cost, long-range unmanned aerial vehicle (UAV) data logger using long range (LoRa) module," in *IEEE Symposium on Wireless Technology and Applications, ISWTA*, Aug. 2022, vol. 2022-Augus, pp. 22–28, doi: 10.1109/ISWTA55313.2022.9942751.

[5]   W. San-Um, P. Lekbunyasin, M. Kodyoo, W. Wongsuwan, J. Makfak, and J. Kerdsri, "A long-range low-power wireless sensor network based on U-LoRa technology for tactical troops tracking systems," in *Proceedings - ACDT 2017: 3rd Asian Conference on Defence Technology: Advance Research Collaboration on Defence Technology*, Jan. 2017, pp. 32–35, doi: 10.1109/ACDT.2017.7886152.

[6]   S. Alabed, "Noncoherent distributed beamforming in decentralized two-way relay networks," *Canadian Journal of Electrical and Computer Engineering*, vol. 43, no. 4, pp. 305–309, 2020, doi: 10.1109/CJECE.2020.2991404.

[7]   S. Alabed, I. Maaz, and M. Al-Rabayah, "Improved Bi-directional three-phase single-relay selection technique for cooperative wireless communications," *Computers, Materials and Continua*, vol. 70, no. 1, pp. 999–1015, 2021, doi: 10.32604/cmc.2022.019758.

[8]   S. J. Alabed, J. M. Paredes, and A. B. Gershman, "A Simple distributed space-time coded strategy for two-way relay channels," *IEEE Transactions on Wireless Communications*, vol. 11, no. 4, pp. 1260–1265, Apr. 2012, doi: 10.1109/TWC.2012.020712.102260.

[9]   A. Alsaraira, O. A. Saraereh, A. Ali, and S. Alabed, "Design of LoRa antenna for wearable medical applications," *IEEE Access*, vol. 11, pp. 23886–23895, 2023, doi: 10.1109/ACCESS.2023.3254916.

[10]   P. Lu, "Design and implementation of coal mine wireless sensor Ad Hoc network based on LoRa," in *2022 3rd International Conference on Information Science, Parallel and Distributed Systems, ISPDS 2022*, Jul. 2022, pp. 54–57, doi: 10.1109/ISPDS56360.2022.9874124.

[11]   S. Mishra, S. Nayak, and R. Yadav, "An energy eficient LoRa-based multi-sensor IoT network for smart sensor agriculture system," in *2023 IEEE Topical Conference on Wireless Sensors and Sensor Networks, WiSNeT 2023*, Jan. 2023, pp. 28–31, doi: 10.1109/WiSNeT56959.2023.10046242.

[12]   A. I. Ali, S. Z. Partal, S. Kepke, and H. P. Partal, "ZigBee and LoRa based wireless sensors for smart environment and IoT applications," in *Proceedings - 2019 IEEE 1st Global Power, Energy and Communication Conference, GPECOM 2019*, Jun. 2019, pp. 19–23, doi: 10.1109/GPECOM.2019.8778505.

[13]   Y. W. Ma and J. L. Chen, "Toward intelligent agriculture service platform with lora-based wireless sensor network," in *Proceedings of 4th IEEE International Conference on Applied System Innovation 2018, ICASI 2018*, Apr. 2018, pp. 204–207, doi: 10.1109/ICASI.2018.8394568.

[14]   S. Yogalakshmi and R. Chakaravathi, "Development of an efficient algorithm in hybrid communication for secure data transmission using LoRa technology," in *Proceedings of the 2020 IEEE International Conference on Communication and Signal Processing, ICCSP 2020*, Jul. 2020, pp. 1628–1632, doi: 10.1109/ICCSP48568.2020.9182233.

[15]   Z. AlArnaout, N. Mostafa, S. Alabed, W. Aly, and A. Shdefat, "A. RAPT: A robust attack path tracing algorithm to mitigate SYN-flood DDoS cyberattacks," *Sensors*, vol. 23, no. 102, pp. 1-20, 2023, doi: 10.3390/s23010102.

[16]   W. T. Sung, S. J. Hsiao, S. Y. Wang, and J. H. Chou, "LoRa-based internet of things secure localization system and application," in *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, Oct. 2019, vol. 2019-October, pp. 1672–1677, doi: 10.1109/SMC.2019.8913875.

[17]   J. Xing, L. Hou, K. Zhang, and K. Zheng, "An improved secure key management scheme for LoRa system," in *International Conference on Communication Technology Proceedings, ICCT*, Oct. 2019, pp. 296–301, doi: 10.1109/ICCT46805.2019.8947215.

[18]   A. Iqbal and T. Iqbal, "Low-cost and secure communication system for remote micro-grids using AES cryptography on ESP32 with LoRa module," in *2018 IEEE Electrical Power and Energy Conference, EPEC 2018*, Oct. 2018, pp. 1–5, doi: 10.1109/EPEC.2018.8598380.

[19]   P. Edward, S. Elzeiny, M. Ashour, and T. Elshabrawy, "On the coexistence of LoRa-and interleaved chirp spreading LoRa-based modulations," in *International Conference on Wireless and Mobile Computing, Networking and Communications*, Oct. 2019, vol. 2019-Octob, pp. 1–6, doi: 10.1109/WiMOB.2019.8923211.

[20]   I. B. F. De Almeida, M. Chafii, A. Nimr, and G. Fettweis, "In-phase and quadrature chirp spread spectrum for IoT communications," in *2020 IEEE Global Communications Conference, GLOBECOM 2020 - Proceedings*, Dec. 2020, vol. 2020-Janua, pp. 1–6, doi: 10.1109/GLOBECOM42002.2020.9348094.

[21]   F. I. Abro, F. Rauf, M. Batool, B. S. C. Dhry, and S. Aslam, "An efficient speech coding technique for secure mobile communications," in *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference, IEMCON 2018*, Nov. 2019, pp. 940–944, doi: 10.1109/IEMCON.2018.8614855.

[22] B. Bessette, R. Salami, R. Lefebvre, and M. Jelinek, "Efficient methods for high quality low bit rate wideband speech coding," in *2002 IEEE Speech Coding Workshop: A Paradigm Shift Toward New Coding Functions for the Broadband Age, SCW 2002 - Proceedings*, 2002, pp. 114–116, doi: 10.1109/SCW.2002.1215742.

[23] J. Hai and E. M. Joo, "Improved linear predictive coding method for speech recognition," in *ICICS-PCM 2003 - Proceedings of the 2003 Joint Conference of the 4th International Conference on Information, Communications and Signal Processing and 4th Pacific-Rim Conference on Multimedia*, 2003, vol. 3, pp. 1614–1618, doi: 10.1109/ICICS.2003.1292740.

[24] T. Moriya, Y. Kamamoto, and N. Harada, "Enhanced lossless coding tools of LPC residual for ITU-T G.711.0," in *2010 Data Compression Conference*, 2010, pp. 546–546, doi: 10.1109/dcc.2010.71.

[25] X. Yu, X. You, X. Liu, and C. Li, "An improved algorithm for residual signal excitation based on LPC 10," in *Proceedings - 2022 7th International Conference on Communication, Image and Signal Processing, CCISP 2022*, Nov. 2022, pp. 318–323, doi: 10.1109/CCISP55629.2022.9974264.

[26] F. A. Muin, T. S. Gunawan, E. M. A. Elsheikh, and M. Kartiwi, "Performance analysis of IEEE 1857.2 lossless audio compression linear predictor algorithm," in *2017 IEEE International Conference on Smart Instrumentation, Measurement and Applications, ICSIMA 2017*, Nov. 2018, vol. 2017-November, pp. 1–6, doi: 10.1109/ICSIMA.2017.8312033.

[27] I. Mansour and S. Al-Abed, "A New Architecture Model for Multi Pulse Linear Predictive Coder for Low-Bit-Rate Speech Coding," *Dirasat: Engineering Sciences*, vol. 33, no. 2, 2010, [Online]. Available: http://journals.ju.edu.jo/DirasatEng/article/view/1839.

[28] I. Mansour and S. Alabed, "Using sinusoidal model to implement sinusoidal speech coder with speech enhancer," in *The 6th Jordanian International Electrical and Electronics Engineering Conference (JIEEEC)*, 2006, pp. 1–8.

[29] I. Mansour, S. AL-Abed, and A. Atieh, "A new technique for regular pulse predictive coding of speech at low bit rates," *The International Conference on Electrical Engineering*, vol. 6, no. 6, pp. 1–13, May 2008, doi: 10.21608/iceeng.2008.34245.

BIOGRAPHIES OF AUTHORS

Samer Alabed ⓘ 🄶 SC 🅓 is currently an Associate Professor and the Head of Biomedical Engineering Department at the German Jordanian University, Jordan. He was an associate professor of Electrical Engineering at the college of engineering and technology in the American University of the Middle East, Kuwait, from 2015 to 2022. He also worked in Darmstadt University of Technology, Darmstadt, Germany, from 2008 to 2015. He received his Ph.D. degree in electrical engineering and information technology from Darmstadt University of Technology, Germany. During the last 18 years, he has worked as an associate professor, assistant professor, researcher, and lecturer in several universities and supervised tens of master students and Ph.D. students. He received several awards and grants from IEE, IEEE, DAAD, DFG, ERC, EU, and AUM. He was invited to many conferences in Europe, US, and North Africa. Further information is available on his homepage: http://drsameralabed.wixsite.com/samer. He can be contacted at email: Samer.Alabed@gju.edu.

Amer Alsaraira ⓘ 🄶 SC 🅓 holds a Ph.D. in biomedical engineering from Monash University since the year 2009. He is currently an assistant professor at Biomedical Engineering Department at German Jordanian University-Jordan since 9/2022. Alsaraira also worked as an assistant professor at the department of electrical engineering at the American University of Middle East (AUM)-Kuwait, for the period between 8/2019-9/2022 and worked as an assistant professor at Biomedical Engineering Department at Hashemite University-Jordan for the period between 12/2009 and 9/2019. He is teaching various courses for the undergraduate students, supervising their graduation projects, supervised undergraduate students, and a member of many committees at the department. He participated in many training workshops to support the university efforts toward gaining ABET accreditation. His current research is in the fields of modeling and simulation of biomedical systems, wireless networks, and DSP. He can be contacted at email: amer.alsaraira@gju.edu.jo.

Nour Mostafa ⓘ 🄶 SC 🅓 received the Ph.D. degree from the Queen's University Belfast School of Electronics, Electrical Engineering and Computer Science, UK, in 2013. He was a Software Developer with Liberty Information Technology, UK. He is currently an Associate Professor of computer science with the College of Engineering and Technology, American University of the Middle East. His current research interests include cloud, fog and IoT computing, grid computing, large database management, artificial intelligence, and distributed computing. He can be contacted at email: nour.moustafa@aum.edu.kw.

Mohammad Al-Rabayah holds a Ph.D. in electrical engineering from the University of New South Wales since 2011. He is currently an assistant professor at the American University of Middle East-Kuwait, since 2018. He is teaching various courses for the undergraduate students, an academic advisor for electrical engineering students, and a member of many committees for maintaining high level of education and research at the department. Al-Rabayah also worked as an assistant professor at Prince Sultan University, Riyadh-KSA, from 8/2015 to 8/2018. During his time, he taught many courses at the communications and networks department and participated to support the university efforts toward gaining ABET accreditation. He worked as an assistant professor at Prince Sattam bin Abdulaziz University, Kharj-KSA, from 10/2014 to 6/2015. His current research is in the fields of wireless networks, and cooperative wireless communications. He can be contacted at email: mohammad.alrabayah@aum.edu.kw.

Yehia Kotb received the Ph.D. degree in field of workflows and process optimization from the Faculty of Computer Science, University of Western Ontario, London, ON, Canada, in 2011. He is currently an Assistant Professor of Computer Engineering with the College of Engineering and Technology, American University of the Middle East (AUM), Al 'Uqiylah, Kuwait. Prior to joining the AUM, he was a Senior Software Developer with the Akira Systems, London, ON, Canada. His research interests include probabilistic process learning and multiagent cooperation in distributed systems. He can be contacted at email: yehia.kotb@aum.edu.kw.

Omar A. Saraereh received the B.S. degree in telecommunication engineering from Mutah University, Jordan, in 1999, the M.Sc. degree in digital communication systems in U.K., and the Ph.D. degree in electrical and electronic engineering from Loughborough University, U.K., in 2005. From 2001 to 2005, he was a member of staff with the Centre for Mobile Communication Research, Loughborough University. He has more than 17 years of academic and practical experience in electrical engineering, mobile communications, various antennas design, fabrication and measurements, radiation hazards and health effects, and wireless communications. He was a high-level Consultant and a Turnkey Solution Originator in countless business and charitable sectors, and an International Public Speaker and a Trainer on a variety of business and people management topics. He is currently a full Professor with the Department of Electrical Engineering, Hashemite University, Jordan. He has published many articles in various international journals and conferences. He can be contacted at email: eloas2@hu.edu.jo.