Chaos Theory and DNA Computation Based Data Encryption System for E-Healthcare Monitoring System

Dr. Ali Fadel Marhoon
University of Basrah-Computer science department

Ali Hussein Hamad
University of Baghdad - Information and communication engineering department

Abstract

The body area sensor network consists of small motes, such devices are considerably energy constrained, limited computation capabilities, and small size of memory such that most advanced encryption scheme cannot be implemented in this type of sensor network. An encryption algorithm must be designed to be a tradeoff between simple computation and powerful encryption scheme. Our proposed algorithm consists of combination of two approaches DNA computation and chaos theory to improve the one-time pad encryption technique. In this work a small amount of memory capacity is required since acquiring samples were encrypted individually in real time. The proposed algorithm appears to be very secure. In this paper, we will explain briefly the design of the proposed algorithm and show its efficiency through encryption tests. Also the effect of some sample loss due to collision occurs in the communication has been studied. The algorithm has been simulated using MATLAB and tested practically using shimmer sensor network platform.

Keywords: chaos theory, DNA computation, Body area sensor network.

1. Introduction

Recent advances in e-healthcare monitoring system, a wireless technology have develop a sensor nodes capable of sensing, processing, and communicate for several human body vital signs such as ElectroCardioGram (ECG), ElectroMyoGraphy (EMG), Galvanic Skin Response (GSR) ,body temperature ,etc. These sensor nodes collaborate to form a body area network BAN. The security of patient's personal health data is a critical issue when these data being transfer on a wireless channels to the end user , (doctors, nurse, family member, or other authorized people) where the medical decisions are made based on the data received. The data exchanged wirelessly where cables can't be used, so all nodes can receive data if they are within range of this wave. If the network is not secure, an adversary could read, modify, and inject messages to the network; this will confuse the end user who makes the decision for the patient's case (wrong decision may lead to a worst case).

A powerful method of data encryption is the one-time-pad algorithm since it is consider being unbreakable [1], where each single piece of data is encrypted alone with a unique key. The disadvantage of this method is how to generate such huge number of keys; a pseudorandom number generator PRNG could be used to generate the required key but it has a problem of key repetition. So a good approach is to use the chaos theory to generate these keys which eliminate the problem of repetition.

DNA cryptography emerged as a new cryptographic field, in which DNA is used as an information carrier and modern biological technology is used as implementation tool [2, 3]. Recent study shows that DNA Computing can be used as a new efficient method to solve difficult mathematical problems [4]. Adleman [4], in 1994 solved Hamiltonian path problem (HPP) by using DNA computing and its advantages such as vast parallelism and extraordinary information density.

The rest of the paper is organized as follows. In section 2 and 3 an overview of DNA computation and chaos theory. In section 4 we review one-time pad encryption basics which are necessary to understand the proposed algorithm. In section 5 a details of the encryption and decryption process based on DNA and chaos theory were presented. In section 6 a description of the key space analysis of the proposed algorithm. Section 7 describes the correlation coefficient analysis for the algorithm. Section 8 describes the hardware and software implementation. Section 9 discusses the results of both, simulation and the real time implementation. Finally a conclusion of the paper and light on future directions were discussed in section 9.

2. DNA computation

DNA stands for **Deoxyribo Nucleic Acid** which is a biochemical macromolecule that contains genetic information necessary for the functioning of living beings. The field of DNA computing was initially developed by Leonard Adleman [4] at the University of Southern California. In 1994, Adleman demonstrated a proof-of-concept use of DNA as a form of computation which solved the seven-point Hamiltonian path problem. DNA is organized as chromosomes in a cell’s nucleus and the chromosomes make up the genome, the entire hereditary information about a cell. A DNA molecule consists of two strands of nucleotides twisted together to form a double helix figure(1). Four kinds of bases are found in the two strands, namely **ADENINE (A)**, **GUANINE (G)**,

THYMINE (**T**), and **CYTOSINE** (**C**). A strand contains a sequence of bases in a specific pattern. The other strand contains the complementary nucleotides of the first strand. Adenine pairs with thymine using a double bond (A = *T)*, while thymine and cytosine pair with each other using a triple bond (*G≡C)*. The genomic sequencing information is contained in the nucleotide bases.

Various operations could be performed on DNA are ligation, polymerase chain reaction (PCR), gel electrophoresis and affinity purification. DNA computing is an inter-disciplinary area concerned with the use of DNA molecules for the implementation of computational processes. The main features of DNA are massive parallelism and complementarity as proposed by Watson and Crick [5].

3. Chaos

Chaos functions have mainly used to develop mathematical models of nonlinear systems. They have attracted the attention of many mathematicians owing to their extremely sensitive nature to initial conditions and their immense applicability to modeling complex problems of daily life. Chaotic functions which were first studied in the 1960's by Lorenzo [6] show numerous interesting properties. The sequences produced by such functions has very good random and complexity. These functions have an extreme sensitiveness to initial conditions. For example, if the initial start value of a chaotic function is modified 10^{-20}, iterative numbers produced after some iterations are completely different from each other. This extreme sensitivity to initial conditions and some other interesting properties, such as pseudo-randomness, wide spectrum and good correlation, grant chaotic functions as a promising alternative for the conventional cryptographic algorithms.

The main advantage using chaos lies in the observation that a chaotic signal looks like noise for the unauthorized users. Moreover, generating chaotic values is often of low cost with simple iterations, which makes it suitable for the construction of stream ciphers. Therefore, cryptosystem can provide a secure and fast means for data encryption, which is crucial for data transmission in many applications. Generally speaking, chaotic stream ciphers use chaotic systems to generate pseudorandom key stream to encrypt the data one by one. In this work, we choose a 1-D discrete logistic map as a chaotic system, where it generates several efficient random sequences. The mathematical discrete form of the 1-D logistic map is given by:

$$f(x_n) = x_{n+1} = rx_n(1 - x_n)$$

Where x_n is the state variable being in the interval [0, 1] , $r \in (3.569945,4), x_0 \in (0,1)$ [7]. The value of x_n is multiply by 65535(16-bit which matching value sensed by sensor board) and rounded to the nearest integer equals or less than x_n.

4. One time pad

The way the one-time pad work is very simple, the process is to encrypt each sample in the data by a modular addition which gets a bit or character from a random key generator. For example, let the key sequence generated by a random generator is:

$pad=k_1k_2....k_n$

where $k_i \in \{0,1\}$

The original message which will be encrypted by the pad keys is:

$message=m_1m_2....m_n$

where $m_i \in \{0,1\}$

then the cipher is:

$c_i=m_i \oplus k_i$

To decrypt the cipher in the receiver side the following function is used:

$m_i=(m_i \oplus k_i) \oplus k_i$

5. Proposed method for sampling data encryption and decryption

In this section, we discuss the detailed of the proposed algorithm for data encryption step by step using DNA encoding technique with chaotic logistic maps as well as decryption process figure(2). The proposed algorithm was tested using the MIT-BIH ECG signal database for simulation and study purposes [8].

Step 1: convert the signal samples into a binary sequence as an $n\times m$ binary matrix.

Step 2: The binary sequence is encoded into a matrix of nucleotides (DNA sequence matrix) to obtain the encoding matrix $n\times m/2$ such that:

A↔ 00, T↔ 01, C↔ 10, G↔ 11.

For example let the signal data array $s = \begin{bmatrix} 2 \\ 7 \\ 10 \end{bmatrix}$, then the binary sequence would be $s = \begin{bmatrix} 00000010 \\ 00000111 \\ 00001010 \end{bmatrix}$

and the DNA sequence is $s = \begin{bmatrix} AAAC \\ AATG \\ AACC \end{bmatrix}$

Step 3: Apply the Polymer Chain Reaction (PCR) which is an algorithm to divide the DNA sequence matrix into 4 sub matrices each $n\times m/4$:

DNA sub matrix1 is $s_1 = \begin{bmatrix} A \\ A \\ A \end{bmatrix}$

DNA sub matrix2 is $s_2 = \begin{bmatrix} A \\ A \\ A \end{bmatrix}$

DNA sub matrix3 is $s_3 = \begin{bmatrix} A \\ T \\ C \end{bmatrix}$

DNA sub matrix4 is $s_4 = \begin{bmatrix} C \\ G \\ C \end{bmatrix}$

Step 4: Generate a chaotic sequence vector $n\times 1$ through 1D logistic map with initial condition x_o with word length equal to m.

Step 5: Apply steps 1 to 3 for the chaotic sequence.

Step 6: Add the DNA sub matrices of the original signal to the DNA sub matrices of the chaotic sequence according to the rules shown in table1 [1,7]:

Step 7: Recombine the sub matrices generated from step 6 to form a new binary sequence matrix c $(n\times m/2)$.

Step 8: generate two chaotic sequence c_1 $(n\times 1)$ and c_2 $(1\times m)$ with another initial condition x_{o2}, x_{o3}, multiply them to produce a matrix w $(n\times m)$. Map the value of w into (0,1) by mod(w,1), then use the following threshold function $f(x)$ to get the binary sequence matrix:

$$f(x) = \begin{cases} 0, & 0 < w(i,j) \leq 0.5 \\ 0, & 0.5 < w(i,j) \leq 1 \end{cases}$$

Step 9 : If $w(i,j)=1$, then $c(i,j)$ is complemented, otherwise it is unchanged. This would produce a new encoding matrix c' $(n\times m/2)$.

Step 10: Apply the inverse process of ***step2*** and ***step 1*** for the sequence matrix c', then we obtain a real value matrix D which represent the encrypted data signal.

In the decryption process, the reverse process is applied from ***step 10*** to ***step 1*** except that instead of addition use the subtraction process as shown in table 2.

6. Key space analysis

The key-space of an encryption technique is the set of possible keys that can be used to encode data using that technique. In the case of a strong encryption scheme, many keys must be tried in any attack on that technique. Our proposed algorithm includes six parameters for generating the chaotic maps (the initial values used to generate the chaotic map) and one is the matrix complement (step 8). Hence the key space of the proposed algorithm includes seven groups of system parameters, which could be calculated by the following equation:

$$Key\ space = k_1 \times k_2 \times k_3 \times k_4 \times k_5 \times k_6 \times k_7$$

7. Correlation analysis

The correlation coefficients is an important features, it is calculated based on correlation between the encrypted signal and the original signal. The main points obtained from the correlation coefficients are mentioned below:

- When it is closed to 1, then there is a positive linear relationship between the two vectors.
- When it is closed to -1, then there is a negative linear relationship between the two vectors.
- When it is closed to 0, then there is no linear relationship between the two vectors.

The following formulas are used to find the correlation coefficient:

$$E(x) = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$D(x) = \frac{1}{N}\sum_{i=1}^{N} (x_i - E(x))^2$$

$$cov(x) = \frac{1}{N}\sum_{i=1}^{N} (x_i - E(x))\,(y_i - E(y))$$

$$r_{xy} = \frac{cov(x,y)}{\sqrt{D(x)} \times \sqrt{D(y)}}$$

8. Practical implementation of the encryption algorithm

In this work, the target embedded sensor system is the SHIMMER platform [9]. From the hardware viewpoint, this platform includes a low-power 16-bit microcontroller, Texas Instrument MSP430F1611, a low-power radio supported with 802.15.4 radio, and an extension module for ECG, EMG, and GSR and built in 3-D accelerometer acquisition. The MSP430 microcontroller runs at 8 MHz, has 10 KB of RAM, 48 KB of flash, and includes a fast hardware multiplier.

In our implementation, five sensor nodes where used as a body area networks, one of these node is used to be a coordinator node (the slowest node), while the encryption algorithm where implemented inside the sensing nodes. The important issue which has a major effect on the performance of the decryption process is the throughput of the system which can be minimizing if a collision has been occurs in the MAC layer. The resulting demand for retransmission deteriorates the throughput; here we study the effect of the collision on the decryption process.

9. Result and discussion

The proposed algorithm was implemented in MATLAB R2013b for simulation and evaluation. Figure (3) shows the original ECG signal which was taken from MIT-BIH repository. Figure (4) shows the encrypted signal according to the encryption algorithm where the signal seems as a noise signal due to the nature of the chaos function used. Figures (5, 6) respectively show the histogram of the original and encrypted signal. Figure (7) shows the decrypted signal when using the same encryption keys which completely the same as the original signal where the correlation coefficients value was 1. Figure (8) shows the decrypted signal when we use one different key (0.100000001 instead of 0.1), this figure shows how the algorithm is strong when using different key for decryption and also shows the sensitivity of the algorithm to a very small change in the key (initial conditions of the chaos logistic map) this in fact due to the sensitivity of the chaos function to changes in the initial conditions. Figure (9) shows real time data (live ECG signal) and the encryption process is good in the presence of some packet loss. Figure (10) is a zoom to a around the 4200th sample where the packets where lost due to collision occur in the body area network communication, the decryption process assume these samples (~20 successive samples) is zero. The fluctuation shown is due to the chaos nature of the decryption process. The correlation coefficients for deferent set of tests in present of packet loss were around 0.98.

10. Conclusion

In this paper, we propose a real time encryption algorithm based on DNA sequence and chaos theory. The use of chaos as a key generator is more powerful than pseudorandom generator. The DNA computation is a good approach used for data encoding. Furthermore, the nature of the one-time pad to encrypt each sample individually makes it appropriate technique to be used in wireless sensor network were it minimizes the required memory space. Finally the algorithm has been implemented successfully in real time body area network for healthcare monitoring system and was very suitable to be used in the presence of collision in wireless communication. For future work, it is important to develop an algorithm to predict the sample lost and to design a collision free MAC (media access control) protocol.

References

1- Majid Babaei, A novel text and image encryption method based on chaos theory and DNA computing, Springer, 2012.

2- Gehani A, LaBean T, Reif J (1999) DNA-based cryptography. In: 5th
annual DIMACS meeting on DNA based computers (DNA 5),MIT, Cambridge, MA, June 1999

3- Grash Jacob, A.Murugan," DNA based cryptography:An overview and analysis", International jornal of Emerging Scince 3(1),36-27, March2013.

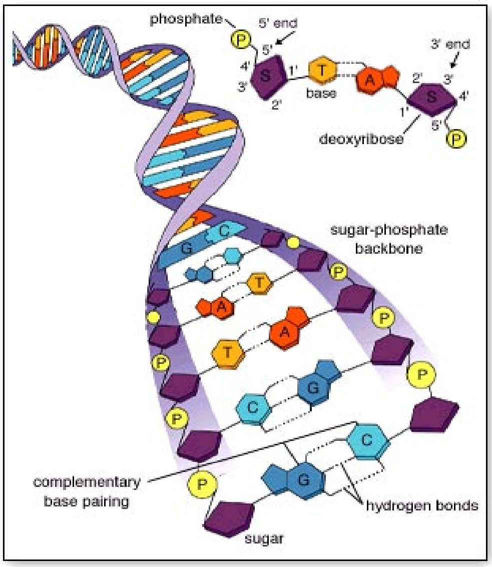
4- Adleman L M, “Molecular Computation of Solutions to Combinatorial Problems, Science”, 266:1021-1024, November 1994.
5- Watson,J.D., Crick,F.H.C., “A Structure for De-oxy Ribose Nucleic Acid”, Nature, vol. 25(1953), pp. 737-738
6- Lorenz, E.N. Deterministic non-periodic flow. J. Atmos. Sci. 20, (1963), 130--141.
7- Qiang Zhang, Ling Guo, Xiaopeng Wei, Image encryption using DNA addition combining with chaotic maps, Mathematical and Computer Modelling 52(2028-2035), Elsever, 2010.
8- MIT-BIH arrhythmia database. (2005). [Online]. Available: http://www. physionet.org/physiobank/database/mitdb/
9- SHIMMER Research, “Shimmer Small Wireless Sensor Platform Designed to Support Wearable Applications.” [Online]. Available: http://shimmer-research.com
10- Pardeep Kumar, Hooxbc n-Jae Lee, Security Issues in Healthcare Applications Using Wireless Medical Sensor Networks: A Survey, 12, 55-91; doi:10.3390/s120100055, Sensors 2012.
11- Lipton R.J., “Using DNA to solve NP-complete problems”, Science, (1995), 268(4), pp. 542-545.

Table 1 Addition operation for DNA sequence

Add	A	T	C	G
A	A	T	C	G
T	T	C	G	A
C	C	G	A	T
G	G	A	T	C

Table 2 Subtraction operation for DNA sequence

Sub	A	T	C	G
A	A	G	C	T
T	T	A	G	C
C	C	T	A	G
G	G	C	T	A



Figure(1) A DNA molecule representation using double-helical structure

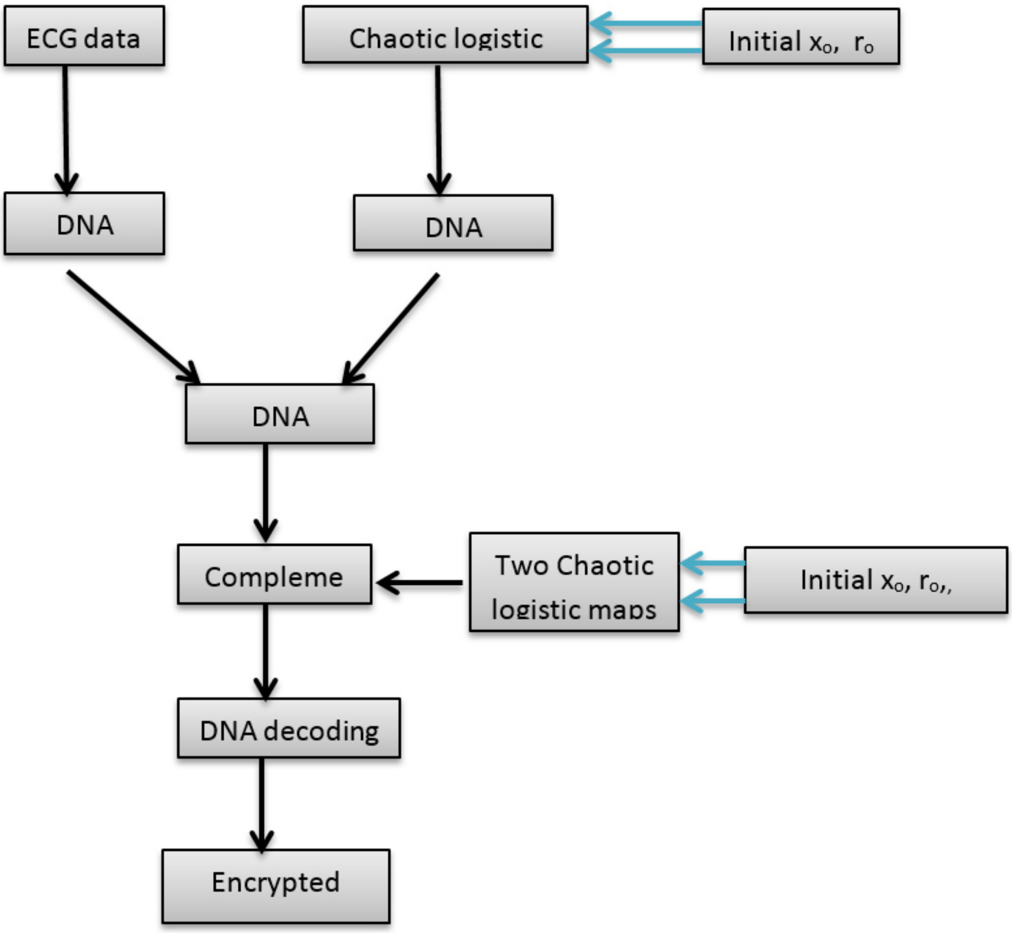


Figure (2) Proposed encryption algorithm based on DNA computing and chaos theory

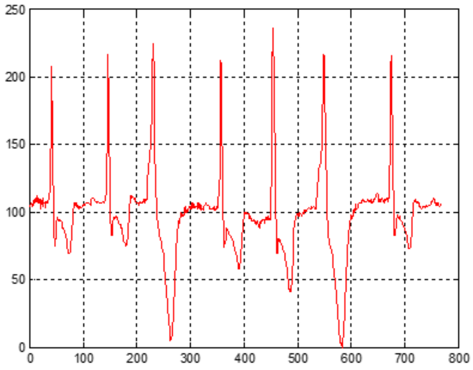


Figure (3) Original ECG signal

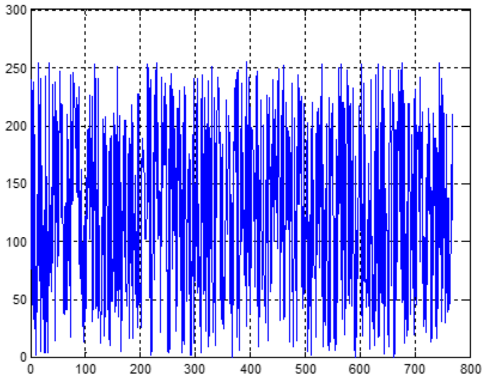


Figure (4) Encrypted ECG signal

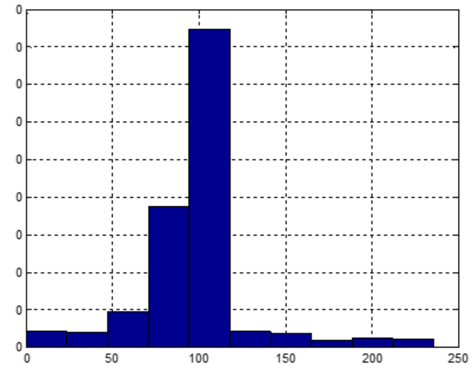


Figure (5) Histogram of the original ECG signal

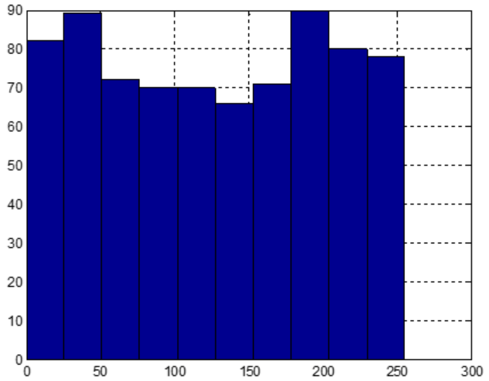


Figure (6) Histogram of the encrypted ECG signal

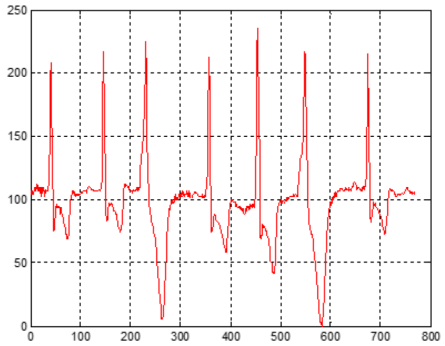


Figure 7 decrypted ECG signal with key value

xo = 0.1,0.2, 0.3, and ro = 3.9

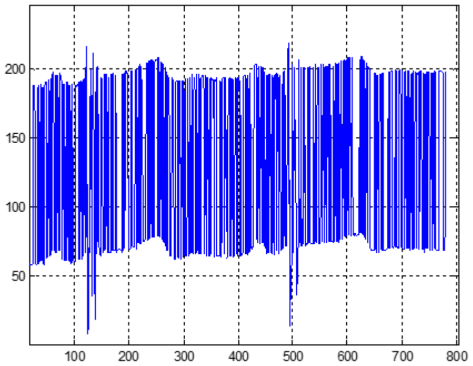


Figure 8 decrypted ECG signal with key value

x_o = 0.100000001,0.2, 0.3, and ro = 3.9

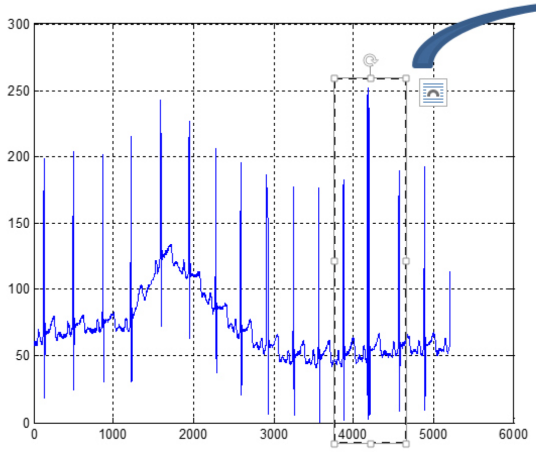


Figure (9) real time ECG signal with some loss in samples

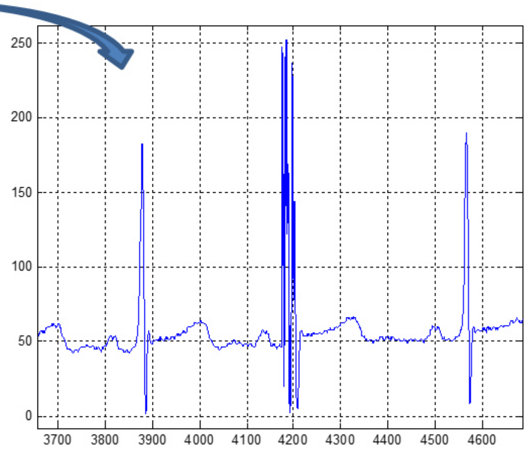


Figure (10) a zoom window near the 4200th sample of figure (9) shows the effect of samples loss