A Lightweight High Capacity ECG Watermark with Protection against Data Loss

Arezou Soltani Panah
Computer Science and Information Technology
RMIT University, Melbourne, Australia
arezou.soltanipanah@rmit.edu.au

Ron van Schyndel
Computer Science and Information Technology
RMIT University, Melbourne, Australia
ron.vanschyndel@rmit.edu.au

ABSTRACT
Wireless Body Sensor Networks are used for pervasive health monitoring and often composed of a large number of sensors communicating over wireless connections. The wireless nature of WBSN poses challenges in terms of security and reliability since sensory data are vulnerable to interception, intrusion and modifications. Moreover, there are ergonomic limitations such as size and weight on wearable sensors that restrict their computational power. In this paper, we propose a variation of the Wong algorithm used in digital watermarking that is suitable for implementation on simple devices with limited processing capabilities and protects data from tampering between point of origin, which could be a wearable sensor, and the point of distribution which might be a Smartphone. At the same time, the watermark has the advantage to detect and accurately localize degraded data and could rapidly recover after data loss.

Categories and Subject Descriptors
E.3 [**Data Encryption**]: *Public key cryptosystems*;
J.3 [Life Sciences] *Medical Information Systems*;

General Terms
Algorithms, Performance, Reliability, Experimentation, Security, Verification.

Keywords
Digital watermarking, electrocardiogram signal, fragile watermark, robust watermark, wireless body sensor network.

1. INTRODUCTION
The emergence of wireless body sensor networks (WBSNs) and communication environments have changed the way a patient interacts with their physician by just wearing a few sensors that can be monitored remotely instead of leaving the patient in residence. A WBSN consists of a collection of small sensors with limited processing abilities that can be used for early detection or maintenance of chronic conditions of a patient such as arrhythmias. However, this solution has limitations in terms of security and any unauthorized access or modification of patient data en route from sensor to collector could have devastating consequences.

PervasiveHealth 2014, May 20-23, Oldenburg, Germany

DOI 10.4108/icst.pervasivehealth.2014.254931

This is because the wireless medium is inherently less secure than wired media and wireless transmissions can be intercepted, altered, or replayed by an adversary. Moreover, one of the fundamental security considerations in Pervasive Computing environment is system availability since failure could expose devastating medical or financial repercussions. Therefore, it is important to achieve graceful degradation in the presence of node compromise or benign node failure [4].

In the past few years, many new techniques based on encryption have been proposed including end-to-end authentication along the path towards the aggregator node or hierarchical peer-to-peer authentication with complicated dynamic key management schemes [10]. While encryption protects the communication channel, it provides no further protection or evidence of tampering after data is decrypted. Moreover, we need to aggregate sensory data from each sensor node and forward them to a nearby device, which could be a Smartphone or PDA, before transmitting data to a remote server. Sometimes, it is useful to obtain the raw sensory data at the aggregator for further processing such as compression or signal enhancement and make this data more efficient and accurate before forwarding it to the destination. However, aggregation cannot easily be done directly on encrypted data and as a result imposes an additional overhead on decryption. Some works propose using privacy homomorphism [4,5] that allows direct computation on encrypted data but research on this approach is at the beginning stage and also has its own limitations (for example supporting only simple operations such as addition and multiplication [9]). To overcome security issues in WBSNs, we propose a digital watermarking scheme based on Wong's algorithm [1] which involves simple mathematical operations that make it suitable for implementation on low-powered sensors with limited computational resources. Specifically, we focus our studies on electrocardiogram (ECG) signals for this paper, but the proposed algorithm can be applied on any sampled streaming data.

Digital watermarking is the process by which a piece of information is embedded within digital content and it can be used for a variety of purposes such as owner identification, proof of ownership, authentication, or copy control. Fragile (or semi-fragile) watermarks are designed for the detection of any modification in the original content, while robust watermarks can survive normal processing likely to occur such as filtering, lossy compression, as well as direct attack by adversaries intent on removing it. In some applications, fragility may be completely irrelevant, or even undesirable. At the other extreme, there are applications in which the watermark must be fragile to most commonly encountered distortion. In dealing with biomedical data, a relatively small distortion could change the diagnosis of a physician. So, a fragile watermark can be used to detect such distortions. In the case of ECG signals, the diagnosis of cardiac

arrhythmias is highly dependent on the timing of PQRS complex wave structures within ECG signals which are the dominant feature of the ECG signals [20], so it is vitally important that an embedded watermark doesn't distort the PQRS complex waves.

In this work, we introduce a fragile watermark, since our goal is integrity verification of ECG signals that not only detects any tampering or distortion, but also determines the location of modifications. Furthermore, the proposed watermark scheme involves logical operations with scaling estimation which are computationally cheap and make it suitable for resource efficient implementations such as simple low-power sensor devices – see section 5. The proposed watermark can coexist with other standard methods used for protecting data through the communication channel at the distribution stage of the process, such as TLS [2] and since it is applied at the source, it prevents having any un-watermarked data in the network.

In common with many, but not all, watermarking algorithms, our scheme can embed a message [3,19]. For this paper, we have used an image as a watermark message, in common with Wong's original algorithm. This is mainly because an image is human-readable and easier to demonstrate in a court of law — even a damaged one can be recognized by a human jury more easily than binary data. However, we could use other binary patterns such as patient information or some other cyclic redundancy check (CRC) or error correcting codes (ECC) depending on the application. Also, our watermark is reversible, which is a useful feature for performing richer processing at aggregation level on raw sensory data to make them more accurate or compressed.

At the same time, the embedded watermark could be used for protection against data loss. One may argue that packet loss or repetition can be easily handled using a transmission protocol such as TCP which allows for retransmission or packet reordering. However in many streaming applications there is not enough time for retransmission especially in data-intensive usages because it increases network latency and congestion. Furthermore, the quality loss of data itself within a packet cannot be detected in this way. The majority of previous studies for validating acquired data in WBSNs, suffer from the fact that sensors themselves could also be contributor to quality degradation [5]. Even if a CRC code was embedded in the data, it can only assert complete correctness but otherwise no assertions can be made about data quality.

Another possibility could be sending a summary of the received data to the server using an RTCP-like back-channel [6]. Similarly, this method consumes bandwidth and introduces a heavy computation overhead while doesn't necessarily capture all data degradation occurred in the original data itself. In summary, the main contributions of this work include the following aspects:

- Design of a low complexity tamper proofing algorithm for one dimensional floating point data such as ECG with a localisation property which is suitable for simple hardware implementation,
- Adapting the Wong algorithm from two dimensional integer data to one dimensional floating point data using a sliding window approach for locally tracking streaming data,
- Providing a variable-capacity watermarking scheme for embedding secret information inside the data. The capacity is directly related to the degree of signal degradation that can be tolerated and is locally controllable.
- Flexibility against data loss by applying adjustable window size,
- Flexibility in payload contents to implement an icon, or logo, patient data, or various in-band data channel error-correction content.

For this paper, we use the MIT-BIH ECG signal database [24] in our figures.

The remainder of the paper is organized as follows. In the next section we review others' works for watermarking medical signals. Then, in Section 3, we demonstrate our watermarking methodology including watermark insertion and watermark detection and we evaluate the proposed algorithm and show our results in Section 4. Section 5 presents validation of the algorithm in detail. Finally we describe our future work in Section 6.

2. Related Work

The majority of previous works dealing with watermarking of biomedical signals used robust watermark techniques. Engin et al. [14] proposed a discrete wavelet based watermarking scheme for ensuring the security of the ECG signals. The authentication information is embedded in the coefficient sequences of ECG segments in frequency domain. However, the data-carrying capacity of the proposed scheme is limited and also is highly susceptible to estimation attack [18]. Similarly, Zheng and Qian [15] investigated a high capacity wavelet-based algorithm by embedding watermark in selected coefficients of high frequency of Haar wavelet transforms based on lifting scheme which correspond to non-QRS complex waves of the original signal to guarantee the restoration of almost undistorted ECG signals. The selection of coefficients shouldn't distort the QRS complex waves of ECG signals. These watermarking schemes impose very high computational complexity on low-power sensors which is undesirable in WBSNs.

In order to balance the energy consumption among sensors in wireless sensor networks, Wei Zhang et al. [9] suggested a distributed robust watermarking scheme in which each sensor appends a part of the whole watermark into the sensory data at a particular time and the sink node is responsible for watermark verification task which implies that a missing watermark part will allow the sink to locate the corresponding sensor nodes for possible attacks. Although the simple operation of the watermark embedding could balance the energy consumption in the network, they describe the watermark detection operation as being relatively complex. Moreover, the need for such watermarks to be synchronized to each other can make their scheme difficult to apply in practice. The paper did not discuss this issue.

There are a few works that studied fragile watermarking of biomedical signals [16,17]. Kong and Feng [16] compared three different fragile watermarking techniques namely Patchwork [11], least-significant bit (LSB) [12], and quantization index modulation (QIM) [13] for integrity verification of ECG signals and showed that the Patchwork method excelled over others in terms of noise resistance. We also embed watermark directly in signal domain but not necessarily in LSB bit-plane; it can be embedded in multiple contiguous bit-planes which make it more resilient to estimation attacks and increases it data-carrying capacity. Moreover, our algorithm is superior to theirs in terms of detecting the exact location of tampered data. Ibaida et al. [17] developed a low complexity fragile watermarking method based on quantization and LSB methods in which patient medical

information are embedded in ECG signals and scaling parameters are used as the key. The security of their algorithm is based on linear transform values that are used for scaling: scalar multiplicand and offset which are selected based on the ECG acquisition device such as the resolution of the analog to digital converter as they claimed. In comparison, we use a hash value of the key that is more difficult to break from a cryptography point of view.

Kozat et al. [18] used a combination of a robust watermark for storing patient data, and a fragile watermark for identifying possible tampering on host biomedical signal. The robust watermark is 'embedded in the frequency domain' using spread spectrum methodology by dividing a sequence into overlapping subsequences, and 'spreading' the same watermark over many frequencies of these subsequences so that the energy in any one subsequence is very small and minimally detectable. The fragile watermark is added on top of the resulting signal containing the robust watermark by applying LSB alternation on specially selected positions of the signal in order to make it resilient to estimation attack. They also used an Error Correction Scheme, (namely Hamming code), in order to make watermark recoverable during transmission over a noisy channel. Their method is not directly scalable to multiple sensors since having multiple data streams and encoding them in a similar fashion will expose them to estimation attack [3]. For example, an attacker could take an average derived from multiple nodes to estimate the original signal, subtract it from the watermarked signal for each sensor, then repeat the process until it converges. He might not get or understand the watermark sufficiently well, but has enough information to generate an interfering signal and so prevent people from trusting the watermark.

Taking all these consideration into account, in our scheme, we adopt a variation of Wong's algorithm [1] for integrity verification of ECG signals.

The chief benefit is that it is *fragile*, *reversible* and *simple to implement*. Moreover, it is *spatial*, not frequency-based and thus damage can in principle be localized to an individual sample (as opposed to a single window), and is amenable to a fallback procedure described in Section 5.

3. Methodology

3.1 System Model

We consider an eHealth monitoring system where sensor nodes are responsible for continuous data acquisition and then forwarding these to an aggregator device such as a Smartphone over a wireless communication channel. Subsequently, the Smartphone performs aggregation on collected sensory data and transmits the aggregated result to a cloud infrastructure where further medical operations may be taken. In this work, we want to secure the whole path from the point of origin where the data has been captured to the final destination point. To achieve this goal, every sensor will perform the following watermark insertion at data acquisition level before forwarding to the Smartphone. At that point, an aggregator or equivalently a Smartphone, could perform the following watermark extraction algorithm in order to find out tamper detection and replace the fragile watermarks with a robust one for onward transmission, or leave the watermark intact, or perform other operations such as summation or compression and then forward the result to the cloud. We can assume that the transmission will be done over a secure encrypted channel such as TLS between Smartphone and cloud server, and possibly even to sensor nodes. As a result, there is no point along the path from sensor to cloud that data is un-watermarked.

The distinction between the data being watermarked and it being encrypted should be emphasized here. They are not equivalent. Encryption only protects the data while it is encrypted – in transit. Once decrypted, all that protection is lost. If the data is watermarked, it goes where the data goes, and only an active attempt at manipulation would remove trace of it. Here, depending on application, the data can be watermarked from source to destination, with no un-watermarked copy available anywhere. With an encrypted data path, this offers maximal protection.

Fig. 1 shows a generalized overview of the multi-layer system architecture. A wireless communication standard such as Zigbee could be employed for inter-WBSN communications that targets low-cost and low-complexity solutions. It also uses AES algorithm with 128-bit keys to perform authentication.

3.2 Watermark Insertion

P. W. Wong [1] proposed a public key algorithm for image verification. After dividing the image into 8×16 blocks, the seven most significant bits of each block are hashed using a hash function such as MD5, then XORed with a chosen binary logo and finally encrypted using a public key cryptographic method such as RSA. The encrypted bit-stream is inserted in the LSBs of the original block. Our variation of the Wong algorithm reflects the different media used. In our case, we are using 1D floating point stream data, instead of static monochrome images. Now, we demonstrate our method for watermark insertion and detection in more detail.

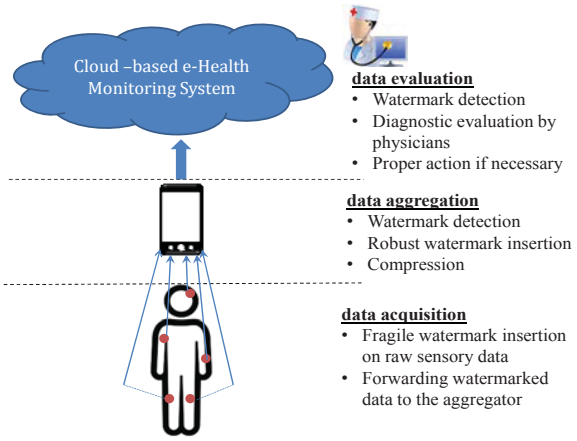

Figure 1: System Model for a Wireless Body Sensor Network

Assume an ECG segment is l seconds long at a sampling rate of r Hz. So there are $n = l \times r$ samples of ECG data — or equivalently an array of n floating point data samples.

To achieve the above in practice, a few pre-processing steps are necessary before inserting the watermark in order to digitize ECGs with m-bit resolution. For so doing, we first scale the floating point data by locally tracking it and finding local minimum and maximum values of that ECG segment within a

window *w*, that is centered on the data value. In other words, a time-based sliding window over the data stream is employed for finding the local minimum and maximum values of that window. As the sliding window proceeds, the minimum and maximum values are updated. Figure 2 shows a window of 16 values (see enlarged part). This explains the quantization ('jumpiness') of the red and blue bounds. After that, the data values are scaled within this range, and rounded to integer intervals.

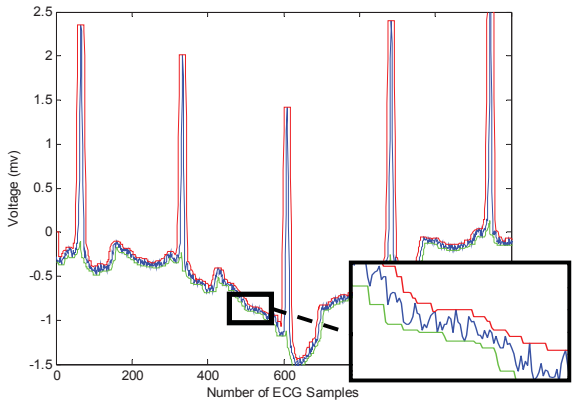

Figure 2: A sliding window tracks data stream incrementally to find local minimum values (green line below signal) and maximum values (red line above signal) for each window.

In contrast to Wong's method that inserted watermark at single bit-planes in each image block, we can pick up multiple contiguous bit-planes for watermark insertion. Masking data with bit-planes, we combine it with the hash value of a given password using an exclusive OR operation. In this work, we used MD5 hash function, but any other cryptographic one-way hash function can be used. As a result, the embedded watermark is dependent on the data which makes it more resistant to copy attack [3] where a watermark can be copied from one content to another. The rationale behind this is based on the fact that even a copied watermark should be combined with another irrelevant carrier, so the inserted watermark will look like noise with respect to the carrier.

The output is then XORed with a binary logo (in the simplest implementation), and we replace bit-planes of the scaled ECG data with the bit-planes of the result of that XOR operation. Finally we undo the scaling by means of the local scaling window parameters. It should be noted that the data is modified in such a way that any watermarked value that was not a local min or max, will not become so after the watermark has been added. This preserves the dynamic scaling and makes it more difficult for an attacker to manipulate the watermark. We only have to defend against fraud, not against removal or rendering the watermark unreadable, since they by definition constitute tampering.

Fig. 3 illustrates the proposed watermarking algorithm. The bit-plane is set to '00011000' which means that 3 most significant bits of the data are preserved after watermark insertion, and 2 bits are used as the watermark while the lower 3 bits are unchanged but subject to round-off error. The result for 8-bit quantization is shown in the figure, but the bit depth is limited only by the natural resolution of the source data (typically 12- or 24-bit ADC). We also need to 'tile' the hash value as well as binary logo to match the size of the original data.

3.3 Watermark Extraction

The extraction procedure involves almost exactly the same process as embedding. Watermarked ECG data are normalized, and the original scaling boundaries are recalculating. Then, normalized ECG data is masked with the bit-planes to form W, XORed with enough tiled hash value of password P. Finally we compute the output block using Logo_est=P⊕W to obtain the logo.

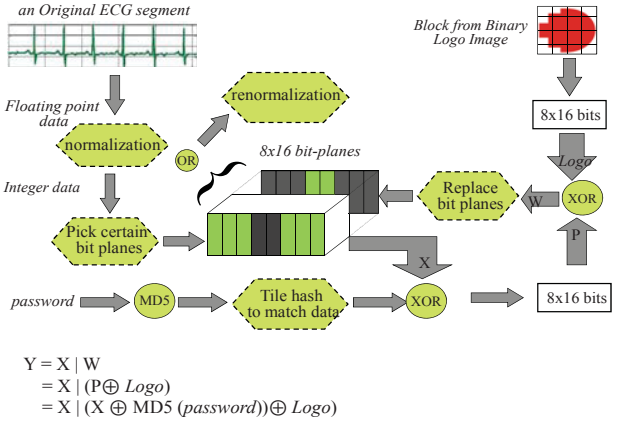

Figure 3. Watermark insertion for bit-plane = '00011000'.

4. Evaluation

4.1 Metrics

One of our difficulties in evaluation of watermarked ECG data or any other non-media data is that we can no longer use the human audio-visual systems as the error measures. For example in the case of audio data, we can use the human audio system to measure the quality of data, or the inaudibility of the watermark using psycho-physical measures. In the case of images we can use human vision to detect the quality of images and visibility of watermark. Here we cannot use any of these human systems to evaluate the quality of watermarked signal, since the diagnostic values of an ECG data is not judged in the same way. We also need an error measure that changes in a same way that the quality of data changes. The majority of recent works used Percentage Residual Difference [15, 17], Signal to Noise Ratio [9,14,16,18], Mean Maximum Error [13] to evaluate the performance of the proposed algorithms.

Giving an unbiased global estimation, these evaluation metrics don't significantly reflect the diagnostic behavior of the watermarked signals. There are a few works that measure the quality of ECG signals by rating the significant of the ECG components. For example, Al-Fahoum [20] proposed a quality measure by decomposition of the signal into sub-bands based on comparing the PQRSTUV complex features of an ECG signal and allocating a weighted score according to diagnostic significance of the sub-bands. However, this requires the use of the original signal. As discussed before, the original data should not appear in any un-watermarked version anywhere. So, we need a metric that measures how sensitive diagnostic error would be. The larger the number, the more likely a faulty diagnosis will result. In other words, we need to determine the quality of data from the data itself. In [18], a cardiologist examines a random subset of watermarked normal and arrhythmia ECG data to identify the diagnostic values and it is claimed that for SNR < 30dB the diagnostic might change because of various distortion near the P-

wave region of ECG signal. However, the proof of this claim should be confirmed.

There are also a few recent works [21,22] that investigate ECG signal quality for false arrhythmia alarm suppression in monitoring eHealth applications using data fusion – filtering, and machine learning techniques to name a few. The existence of these works clearly shows that the quality of ECG signals is not a straightforward task to evaluate and cannot be determined simply by the above metrics.

To be consistent with other works however, we still use PRD and SNR in order to make a direct comparison between watermarking algorithms for comparing the similarity between signals before and after processing and calculate these as follows [23]:

$$PRD = \sqrt{\frac{\sum_{n=1}^{N}[x(n) - x_w(n)]^2}{\sum_{n=1}^{N} x^2(n)}} \times 100 \quad (1)$$

and

$$SNR = 10\, log \frac{\sum_{i=1}^{n} x^2(n)}{\sum_{n=1}^{N}[x_w(n) - x(n)]^2} \quad (2)$$

where N represent the total number of samples in the ECG signal namely $x(n)$ and $x_w(n)$ shows the watermarked ECG signal.

4.2 Models

The process of normalizing and quantizing can be modeled as:

$$d'_{iw} = |d_i|_s + 2^{-b} s M_i \quad (3)$$

$$s = f\left(\max\left(d_{i-\frac{w}{2}}, \ldots, d_{i+\frac{w}{2}}\right), \min\left(d_{i-\frac{w}{2}}, \ldots, d_{i+\frac{w}{2}}\right)\right) \quad (4)$$

where d'_{iw} is the watermarked data d_i for a given window size, w, watermark bits, M_i which encodes b adjacent bits, and scale factor, s. $|d_i|_s$ is the data quantized to a degree corresponding to the maximum and minimum values of the data within the window, and the scale factor, which determines how many bits per sample of the watermark can be encoded. This means that $d_i - |d_i|_s$ represents the maximum error bound that can occur as a result of watermarking. From this point, it is straightforward to use some of the error bounds listed in the references. These limits can be hard coded within the sensor and will thus avoid unbounded error.

For this paper, the function $f(\,)$ in equation (4) is simply taken to be the interval from minimum to maximum, but the form of equation (3,4) leaves open the possibility of making the embedding locally adaptive to some '*diagnostic visibility*' function. The only additional requirement is that $f(\,)$ is invariant to embedding, so that the same $f(\,)$ can be used for extraction. It is self-evident that the maximal error bounds, s, exceed the average PRD for the window at all times, so we have a convenient bound available to the sensor device itself, depending only on the chosen settings, and local extrema.

The encryption and decryption parameters for watermarked data stream D_w can thus be summarized as:

$$D_w = E(D, password, logo, b, m, w) \quad (5)$$

$$Logo_{est} = D(D_w, password, b, m, w) \quad (6)$$

where D is the input data, D_w is the watermarked data, *password* and *logo* are bit arrays, b is the number of bit-planes of the watermark, m is the resolution (in bits) of the full-scale data, and w is the scaling window size. In decoding, we will use the same parameters to form an estimate of the logo, $Logo_{est}$.

4.3 Results

We utilized the 24 hour recorded ECG signals from MIT-BIH database [24] which includes annotated normal and arrhythmia signals with sampling rate of 360Hz, 11 bits of ADC resolution, and 200 adu/mV gain. Table 1 shows the average PRD and SNR for different number of bit-planes. For 11 bit resolution per sample, up to 5 bits per sample can be used for watermarking. We also used a corresponding tiled version of RMIT logo with 20x20 bits = 50 bytes as watermark information.

Table 1. Average PRD / SNR for different watermark capacity where sliding window size=8 and ADC[1] resolution=11 bits.

WM capacity	Normal		Arrhythmia	
	PRD %	SNR (dB)	PRD %	SNR (dB)
1	0.1596	55.4467	0.1012	60.8579
2	0.3425	49.3624	0.2701	51.4602
3	0.7237	42.8647	0.5594	45.9379
4	1.4707	36.9891	1.1547	38.8110
5	2.6144	31.5689	1.7113	35.7312

Figures 4, 5 show how PRD and SNR values vary for inserting watermarking in more significant bit-planes. If watermark is bound to the most significant bits of the signal, the PRD is higher and equivalently SNR is lower and as a result higher diagnostic deficiency. The straight line in Figure 5 demonstrates that the error is directly proportional to the *relative* change in the signal introduced by the watermark.

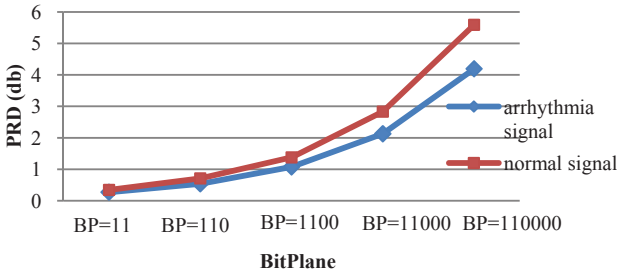

Figure 4. PRD variation for shifting bit-plane to more significant bit-planes of ECG signals.

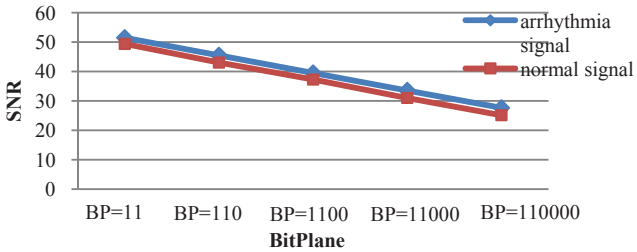

Figure 5: SNR variation for shifting bit-plane to more significant bit-planes of ECG signals.

[1] Analogue to Digital Convertor

5. Verification of Algorithm Properties

In this section, we assert the watermarking properties that we claimed above.

5.1 Computational Complexity

As it is shown in Fig. 3, the Encoding/Decoding process involves three main steps: normalization, hash function calculation, and 3 logical operations (1 AND operation for masking bit-planes and 2 XOR operations). The normalization step includes scaling and binary to decimal (decimal to binary) conversions. As discussed in Section 3.2, the scaling process is done by locally calculating the maxima and minima over a local window. Since it is a standard sliding window, the complexity of windowing process is linear i.e. O(l) where l is the window size because it requires 2 comparisons for each l value in order to find the local maximum and minimum over that window. We can further improve the accuracy of the windowing process by using an overlapping sliding window. As the sliding windows overlap, samples are likely to be considered multiple times, so the boundaries move smoother. A larger overlap will result faster execution but coarser boundaries However, the selected window size here is less than 1/20th sec – an icon appears roughly every second (roughly equal to one heart-beat duration), which doesn't affect the complexity significantly.

As a hash function, we used a standard MD5 which processes the input in 512-bit blocks by executing 64 steps. Each step consists of 4 additions, 3 logical operations, 2 table lookups and 1 rotation. The input message may be arbitrarily large, but since here we just need to calculate the hash value of a password once, it doesn't impact the complexity. Logical operations are also cheap computationally and don't increase the complexity. In Figure 6 the average complexity of encoding and decoding process is shown. The cost of hash calculation and decimal to binary conversions can be considered negligible. The most complex part involves the binary to decimal conversion and other operations including tiling and other necessary array operations.

The total execution time of the encoding and decoding process for an ECG signal including 3600 samples with different watermark capacity is investigated in Table 2. All the experiments were performed on a desktop computer that was equipped with an Intel Core i3 550 CPU (3.20 GHz) and 4GB of RAM. The operating system was a 64-bit version of Microsoft Windows 7.

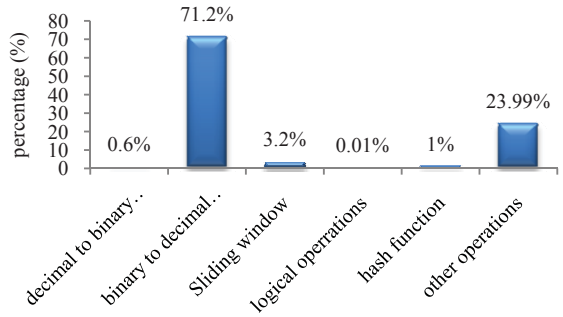

Figure 6: Average Encoding/Decoding Complexity

5.2 Capacity Estimation

The estimated capacity is the upper bound of the number of bits that can be embedded in the host ECG signal without causing diagnostic artifacts. We show this threshold value by Δ which lower bound is proportional to quantization level introduced by the digitizing process.

Table 2: Average execution time for an ECG signal with 3600 samples

WM capacity	Execution time (s)	
	Encoding	Decoding
1	0.296	0.259
2	0.292	0.255
3	0.285	0.271
4	0.306	0.283
5	0.307	0.288

Therefore, the data hiding capacity $C(f)$ for a sampling frequency f is equal to:

$$C(f) = \frac{t \times f}{\Delta} \quad (7)$$

where $\Delta = \left\lfloor \frac{ADC\ resolution}{watermark\ capacity} \right\rfloor$ and t is the total signal time in seconds. Based upon results achieved in Table 1, Δ=[11/5]=2 and as a consequence the total capacity for an ECG signal with ADC resolution of 11 bits and sampling rate of 360 Hz is roughly 1.8 Kbits/sec. With such high capacity, one can easily embed important information inside the ECG signal itself without degrading the signal quality.

5.3 Watermark Graceful Degradation

We also studied how the proposed scheme supports graceful degradation or equivalently how the watermark behaves in the presence of distortion. Fig. 7 shows the RMIT Logo used as watermark information, and segments of the original ECG signal with the logo as input, and the watermarked signal as well as the extracted logo images as output. The density of the logo 'frames' within the signal window depends on how many bits per pixel are stored and on the size of the logo. Fig. 8 shows the effect on the logos of changing individual samples (150 in this case, resulting in 150 x 2bits / sample = 300 entries in error, or 300/20=15 columns), and of deleting those samples. Fig 8 (c) and (d) demonstrates how the watermark behaves in the presence of distortion. Deletions and additions will cause synchronization loss for the remainder of the segment, but it is later reestablished via the protocol.

5.4 Embedding Data Instead of Logo

As mentioned before, we have used an RMIT logo for the embedded data for demonstration purposes. The frame size used in Figures 6, 7, 8 for the logo, is an attribute of the embedded data, not the watermark embedding method. We are not restricted to only using logos, though they are good demonstrators in a court of law. We could instead insert patient data (as long as it contains an error-check of some kind), or other identity, such as the measurement settings and circumstances. A more interesting use is data-chaining, where a low-resolution version of a data frame is stored as watermark within the next frame. If we use a CDMA-encoded binary message with ECC, then synchronization is trivially added, with no extra effort on the part of the collector. Importantly, if a block-chaining method is used, we can vary the message content from frame to frame. The system has the capacity and flexibility for many other such uses.

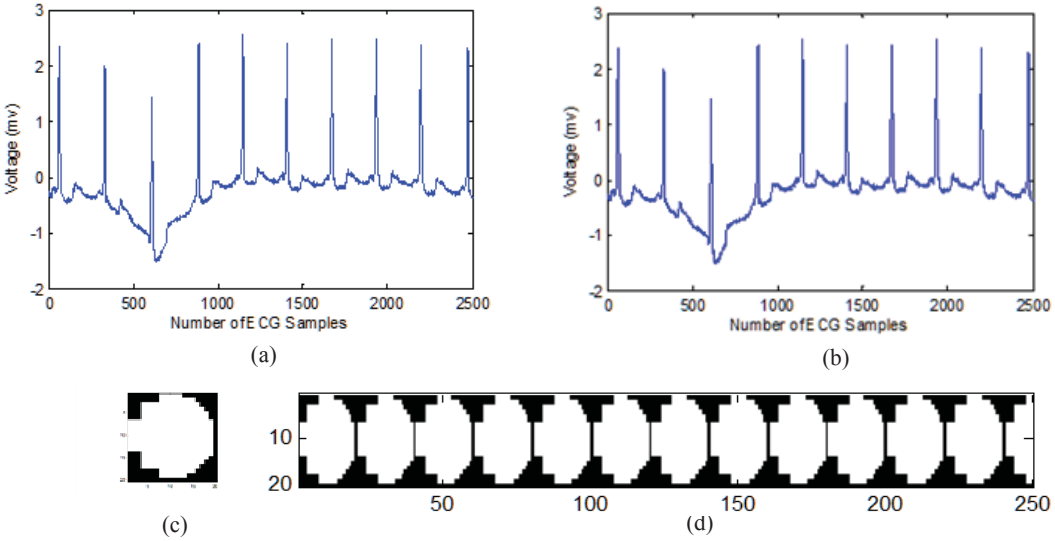

Figure 7. (a,b) Original and watermarked signal, (c) Original 20x20 Logo, (d) Extracted logos

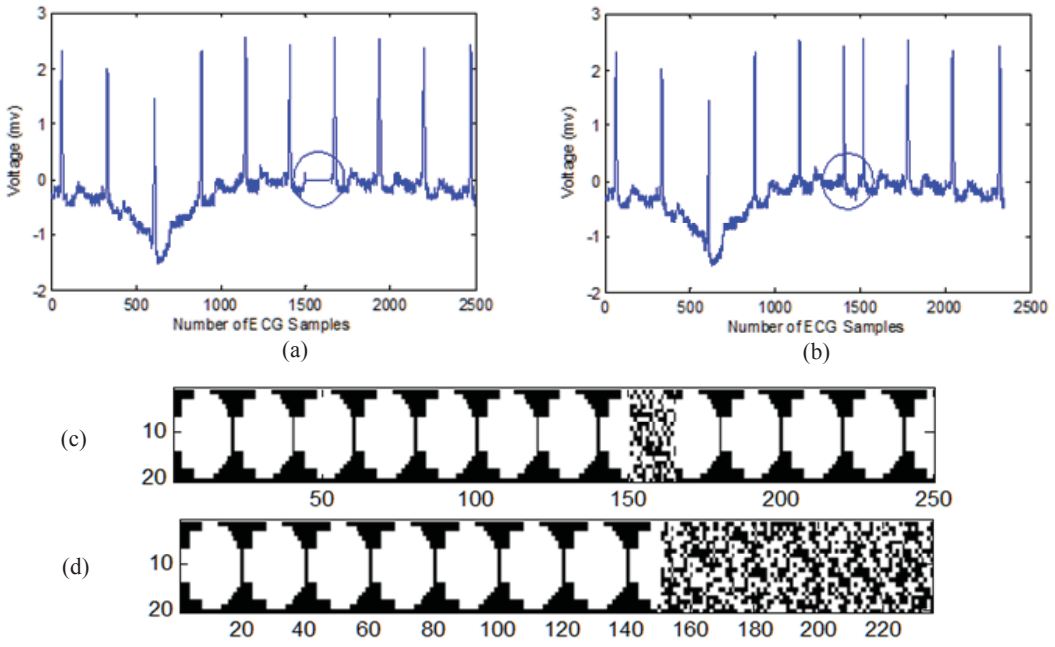

Figure 8. (a) Signal with 0's, (b) Signal with deletions, (c,d) Extracted logos.

6. Conclusion

In this work, we proposed a fragile watermarking algorithm for protection of ECG data at the sensory nodes. Most of the current works are related to the watermark robustness and end-to-end transmission of biomedical data that leaves data vulnerable between sensors and the aggregator where such a watermark is inserted. The main contribution of our work is that it tries to insert a watermark one step after data collection. The advantage of this scheme is its simplicity and ease of implementation in simple devices such as sensors. Evaluation results confirmed that the proposed algorithm is feasible for high accuracy and low complexity for ECG protection. The embedded watermark can be used for tamper proofing because even small distortions in the watermarked data render the watermark non-recoverable. At the same time, our method has the ability to rapidly discover loss of synchronization and search for resynchronization due to applying small windows. If the embedded data uses data link chaining, then loss can be localized. We embedded an icon for the evaluation, but any other binary data can be used as well.

At the aggregation level it might be desirable to remove fragile watermarks and embed a robust one or embed multiple watermarks. As Mintzer [19] suggests "embedding a fragile watermark followed by embedding a robust watermark is bound

to damage the fragile watermark". So, the most robust watermark should be embedded first, and the most fragile watermark should be embedded at last.

Watermarking performance evaluation remains still a challenging problem. However, our method is general and not only applicable to ECG signals, but in the case of ECG, the appropriate parameter would be PRD which is a general measure of damage. In our scheme, any significant local minima/maxima (beyond window size) will remain after embedding and the position of such maxima is unchanged by design. Hence PQRSTUV complex data maxima are maintained in terms of timing. By adapting the window size to be close to detector resolution, we will not change the peak heights significantly either. Therefore, the PRSTUV points themselves are completely recoverable within sensor resolution and as a result the diagnostic significant is relatively unaffected and is limited by PRD.

7. ACKNOWLEDGMENTS

The result of this research has been used to develop an Android Smartphone application. This separate project was funded by the Victorian Government, Department of Business Innovations as Technology Student Accelerator Voucher TVP-TSA-001A. The authors would like to thank Mahathir Almashor and Scientific Technology Pty Ltd, for their collaboration on this project.

8. REFERENCES

[1] Wong, P.W. 1998. A Public Key Watermark for Image Verification and Authentication. In *Proc. of Image Processing* (Chicago, IL, Oct. 04 - 07, 2003), 455-459.

[2] Tirkel, A. Z., Hall, T. E. 2000. Watermarking at point of origin. In *Proc. of ACM Workshop on Multimedia*. MULTIMEDIA '00. ACM, NY, 135-138.

[3] Voloshynovskiy, S., Pereira, S., Pun, T., Eggers, J.J., Su, J.K. 2001. Attacks on Digital Watermarks: classification, estimation based attacks, and benchmarks. *IEEE Communication Magazine*. 39, 8, 118-126.

[4] Xiao, Y. 2007. Security in Distributed, Grid, Mobile, and Pervasive Computing. CRC Press, Auerback Publication.

[5] Andreoni, G., Fanelli, A., Witkowska, I., Perego, P., Fusca, M., Mazzola, M., and Signorini, M. G. 2013. Sensor validation for wearable monitoring system in ambulatory monitoring: application to textile electrodes. *In Proc. of the 7th Int. Conf. on Pervasive Computing Technologies for Healthcare* (Venice, Italy). PervasiveHealth '13. ICST, Belgium,169-175.

[6] Holliman, M. J., Yeung, M. M. 2003. Measurement of data degradation using watermarks, U.S. Patent 0115504 A1 (filing date Dec 19 2001, and issued Jun 2003).

[7] Rodhea, I., Rohner, C. 2008. n-LDA: n-layers data aggregation in sensor networks. In *Proc. of 28th Int. Conf. on Distributed Computing Systems Workshops* (Beijing, June 17 - 20, 2008), 400-405.

[8] Ozdemir, S. 2007. Concealed data aggregation in heterogeneous sensor networks using privacy homomorphism. *In Proc. of 7th IEEE Int. Conf. on Pervasive Services* (Istanbul, July 15 – 20, 2007), 165-168.

[9] Zhang, W., Liu, Y., Das, S.K., De, P. 2008. Secure Data Aggregation in Wireless Sensor Networks: a watermark based authentication supportive approach. *Journal of Elsevier Pervasive Mobile Computing*. 4, 5, 658-680.

[10] Zhang, J., Shankaran, R., Orgun, M., Sattar, A., Varadharajan, V. 2010. A Dynamic Authentication Scheme for Hierarchical Wireless Sensor Networks. In *Proc. of 7th Int. ICST Conf. on Mobile and Ubiquitous Systems (*Australia, Sydney), Springer Berlin, 73, 186-197.

[11] Bender, W., Gruhl, D., Morimoto, N., Lu, A. 1996. Techniques for data hiding. *IBM Systems Journal*, 35(3&4), 313-336.

[12] van Schyndel, R., Tirkel, A. Z., Osborne, C.F. 1994. A digital watermark. In *Proceedings of IEEE Int. Conf. on Image Processing* (Austin, Texas), 86-90.

[13] Chen, B., Wornell, G. 1998. Digital watermarking and information embedding using dither modulation. In *IEEE Second Workshop on Multimedia Signal Processing* (Redondo Beach, CA, Dec. 07-09, 1998) , 273-278.

[14] Engin, M., Cidam, O., Engin, E. Z. 2005. Wavelet transformation based watermarking technique for human electrocardiogram. *Journal of Med Systems*, 29, 6, 589-594.

[15] Zheng K., Qian, X. 2008. Reversible data hiding for electrocardiogram signal based on wavelet transforms. In *Proc. of Computational Intelligence and Security* (Suzhou), 295-299.

[16] Kong, X., Feng, R. 2001. Watermarking medical signals for telemedicine. *IEEE Trans. on Information Technology in Biomedicine*. 5, 3, 195-201.

[17] Ibaida, A., Khalil, I., van Schyndel, R. 2011. A low complexity high capacity ECG signal watermark for wearable sensor-net health monitoring system. In *Proc. of Computing in Cardiology* (Hangzhou), 393-396.

[18] Kozat, S.S, Vlachos, M., Lucches, C., Van Herle, H., Yu, P.S. 2008. Embedding and Retrieving Private Metadata in Electrocardiograms. *Journal of Med Systems*. 33, 4, 241-259.

[19] Mintzer, F., Braudaway, Tom G.W. 1999. If one watermark is good, are more better?. In *Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Phoenix, AZ), 2067-2069.

[20] Al-Fahoum, A.S. 2006. Quality Assessment of ECG Compression Techniques Using a Wavelet-based Diagnostic Measure. *IEEE Trans. on Information Technology Biomedical*, 10, 1, 182-191.

[21] Aboukhalil, A., Nielsen, L., Saeed, M., Mark, R. G., Clifford, G. D. 2008. Reducing false alarm rates for critical arrhythmias using the arterial blood pressure waveform. *Journal of Biomedical Information*, 41, 3, 442-45.

[22] Behar, J., Oster, J., Li, Q., Clifford, G. D. 2013. ECG signal quality during Arrhythmia and Its Application to False Alarm Reduction. In *IEEE Trans. on Information Technology in Biomedical Engineering*, 60,6, 1660–1666.

[23] Cetin, E., Koymen, H. 2000. *Compression of Digital Biomedical Signals*. Boca Raton, Florida.

[24] MIT-BIH Arrhythmia Database, http://www.physionet.org/physiobank/database/mitdb.