Steganography, BioPatterns and Independent Components

David Lowe, B.R.Matam, B.Toch

Neural Computing Research Group, Aston University, Birmingham. B4 7ET, UK.

Abstract

In this paper we discuss the development of steganography in the context of the electronic Patient Health Record. We exploit the signal processing aspects of a frame-based approach of expanding signals using a nonorthogonal basis of independent components derived from the data, and investigate the consequences of applying the framework to different dimensionality biopatterns. The integrity of the hidden message under different benign 'attacks' is investigated along with the capacity of the embedded message. We will particularly concentrate on hiding information in biopatterns such as single channel EEG time series, since this is especially challenging in low dimensional problems such as single channel time series - because information capacity and hence exploitable information redundancy is lower. We will conclude by investigating the additional security of the embedded message by the sensitivity of recovered messages to slight variations in the structure of the independent components, or knowledge of which components have been modified - a topic not previously considered, and which augments any cryptographic approach for security.

1 Introduction

Steganography (Cox, Miller & Bloom, 2002) is the science of hiding information in plain sight. Figure 1 depicts the steganographic process of data watermarking. There is a growing demand in generic digital media for signal processing techniques for hiding information messages inside covertext, to act as digital fingerprinting for applications such as secure data authentication and copyright protection. Digital watermarking may be robust (to resist attacks) or fragile (to indicate whether a digital medium has been compromised). In this work we focus on the issues of blind data authentication for the future electronic patient health record (EPHR). ['blind', since we assume the unmodified source covertext is not available at the receiver.] The future EPHR will need to involve multimodal digital data, and will integrate data of different dimensionalities such as time series, images, freeform and structured text, and video. When a patient or clinician wants to interrogate a person's bioprofile (the integration of this distributed data), there are at least two primary concerns: How to confirm that the retrieved data really does belong to a specific individual (data authentication), and who should have access to that data (data ownership)? Current approaches and proposed standards to the security aspects are concentrating on cryptographically secure transmission. However, having separate tags with this information to the biopatterns is intrinsically open to attack or manipulation and potentially can be lost or modified as part of normal use of the bioprofile.

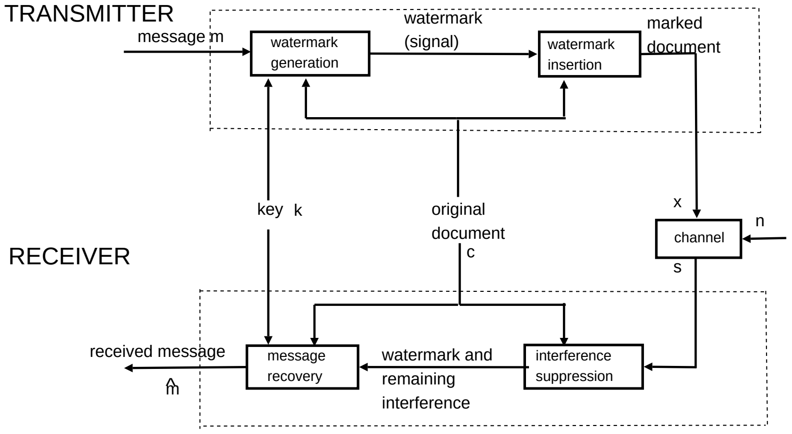


FIGURE 1: Schematic of the digital watermarking process as part of a communication process

Embedding patient metadata inside their biopatterns would be a possible solution, if data rate could be high enough, but not distort clinical information, and also be robust to 'attacks' such as data compression and communication channel distortion. In previous work (Toch & Lowe, 2005; S.Bounkong, B.Toch, D.Saad & D.Lowe, 2003; Bounkong, Saad & Lowe, 2002) we have already investigated the generic use of domain-versatile approaches to steganography. One method we have recently proposed is based on Independent Components. There are two reasons why using independent components is favourable for steganography: the first is that the same framework may be applied to N-dimensional covertexts (the medical data), and so the same approach can be applied across different data domains; The second is that in the context of watermarking, ICA allows the maximization of the information content and minimization of the induced distortion by decomposing the covertext into statistically independent sources. Embedding information across independent sources minimises the emerging cross-channel interference. In fact, for a broad class of attacks and capacity, one can show that distortion is minimised when the message is embedded in statistically independent sources (S.Bounkong et al., 2003). Information theoretical analysis also shows that the information hiding capacity of statistically independent sources is maximal (Moulin & O'Sullivan, 2001). Hence ICA is a form of optimality criterion for steganography in some circumstances.

2 Methodology

The intention is to embed a random binary message m in a feature space of the biomedical covertext $\boldsymbol{C}$ constructed from a basis expansion of $\boldsymbol{C}$. We choose a nonorthogonal independent source model such that $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{S} + \boldsymbol{N}$ where $\boldsymbol{A}$ is a mixing matrix, $\boldsymbol{S}$ are a set of independent sources, and $\boldsymbol{N}$ is a gaussian random noise process. The sources are modified $S \rightarrow \hat{\boldsymbol{S}}$ by an embedding function, $E[\boldsymbol{W}\boldsymbol{C}, \boldsymbol{m}]$ which embeds the random message in the independent sources. We will use an embedding method known as Quantisation Index Modulation (QIM) (Chen & Wornell, 2001), but other embedding approaches could be employed.

The watermarked covertext is then given by

$$\hat{\boldsymbol{C}} = \boldsymbol{A}E[\boldsymbol{W}\boldsymbol{C}, \boldsymbol{m}] + \ker(\boldsymbol{W}) \cap \boldsymbol{C}$$

where $\ker(\boldsymbol{W})$ is the kernel of the linear transformation $\boldsymbol{W}$. We evaluate the demixing matrix on a training set of data representative of the problem and is then held fixed for other similar biopatterns. This precomputed demixing matrix is then communicated to the decoder and acts as a secret key.

The fidelity of the covertext is measured by the distortion $\delta(\boldsymbol{C}, \hat{\boldsymbol{C}})$ between the original and the watermarked biopattern. Clearly this is driven by the data rate of the embedded message, and for clinical applications it is important that the distortion is small and does not interfere with clinical diagnosis. There is thus a tradeoff between message data rate and fidelity. In the case of lossless transmission, this information rate is upper bounded by the channel capacity.

2.1 Decoding

In practice, a test signal $\boldsymbol{c}$ is watermarked using this precomputed demixing matrix by forming $\boldsymbol{s} = \boldsymbol{W}\boldsymbol{c}$ and watermarked to $\hat{\boldsymbol{c}} = A\hat{\boldsymbol{s}}$.

Applying $\boldsymbol{W}$ to the transmitted, watermarked test biopattern, $\hat{\boldsymbol{c}}$ would recover the *watermarked* sources, $\hat{\boldsymbol{s}}$. Using the (known) embedding method (which could also have an optional key to increase security), the watermarked sources could then be decoupled from the estimated message, i.e. $\hat{\boldsymbol{m}} = E^{-1}[\hat{\boldsymbol{s}}]$.

For biomedical applications we are seeking *robustness* of the steganography mechanism: i.e. a watermark should be retrievable from the transmitted data even if the latter has been corrupted by some levels of transmission or processing noise.

2.2 Resilience to non-authorised access

How close can an attacker get to recovering the method, given she knows the method and key factors such as segmentation blocking?

An attacker who intercepts $\hat{\boldsymbol{c}}$ can use the knowledge in $\boldsymbol{c}$ combined with their own knowledge to derive their own estimators of the demixing matrix $\boldsymbol{W}' \neq \boldsymbol{W}$ and watermarked sources $\hat{\boldsymbol{s}}' = \boldsymbol{W}'\hat{\boldsymbol{c}}$. Assuming she knows the embedding mechanism, she constructs an estimate of the message, but using her estimates of the sources, $\boldsymbol{m}' = E^{-1}[\boldsymbol{W}'\hat{\boldsymbol{c}}]$.

3 Example

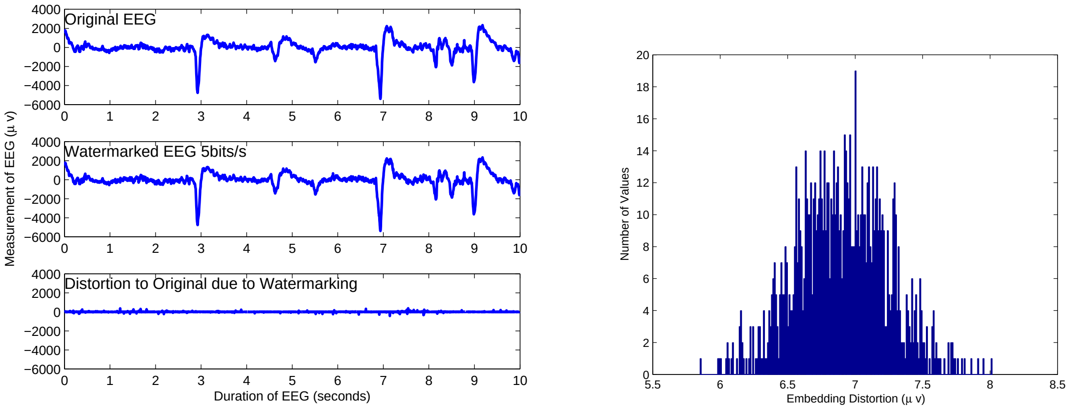
We consider an example of a single channel of EEG time series sampled at 250Hz to 16 bits resolution. We take 30 second segments and divide into one second intervals. Thus $\boldsymbol{C} \in \mathbb{R}^{250\times 30}$, $\boldsymbol{W} \in \mathbb{R}^{30\times 30}$ and there are 30 sources, in this case each of length 250. We are seeking to embed a few bits per second in such medical data to be of practical significance. We use QIM to embed an n bit random message into the sources. We demonstrate the effects of message embedding rate on fidelity, robustness to an 'attack' (quantisation compression of the transmitted EEG), and sensitivity of the message to discovery by a third party.

3.1 Fidelity

The fidelity remains high for message embedding data rates of a few bits per second, and this translates to an insignificant perceptual change of the observed transmitted signal when reviewed by the clinician.

For example, Figure 2(a) shows one example of the distortion induced at $5bs^{-1}$ verifying the lack of perceptual change. Maximum distortion equates to a few microvolts on this scale. Typical magnitudes of recorded EEG are in the range of tens–hundreds of microvolts. Since the embedding process uses random selection of which samples to modify in the sources, this distortion is likely to be a function of where the embedding is applied.

Hence Figure 2(b) depicts a histogram showing the distribution of the *mean absolute deviation* induced by running the steganography process one thousand times. Again, this figure confirms that the distortion over a wide range of embeddings is restricted to a few microvolts.



(a) Comparison of original, watermarked and SnR distortion for a message embedding rate of 5 bits per second

(b) Histogram of the distribution of maximum distortions induced using a sample rate of $5bs^{-1}$ over many random realisations of the message embedding process.

FIGURE 2: Distortion in the original data due to the embedding message at $5bs^{-1}$

3.2 Sensitivity of Reconstructed Message

If an attacker attempts to reconstruct the independent sources and hence extract the private message embedded in the data by the process described previously, how close can they get? We measure similarity between sources using mutual information. For a given estimated source $\boldsymbol{s}_i'$, the most similar original source is identified by seeking the maximum of the mutual information between all sources. Figure 3 plots the maximum mutual information obtained between the extracted sources using the estimated demixing matrix, and the actual demixing matrix. It is clear that the estimated sources are far apart from the true sources.

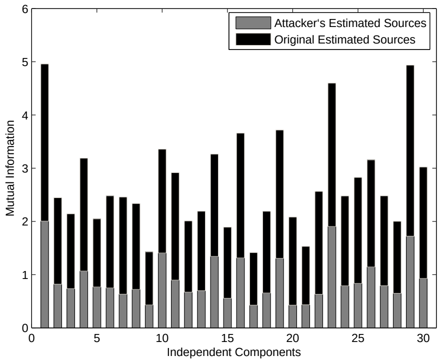
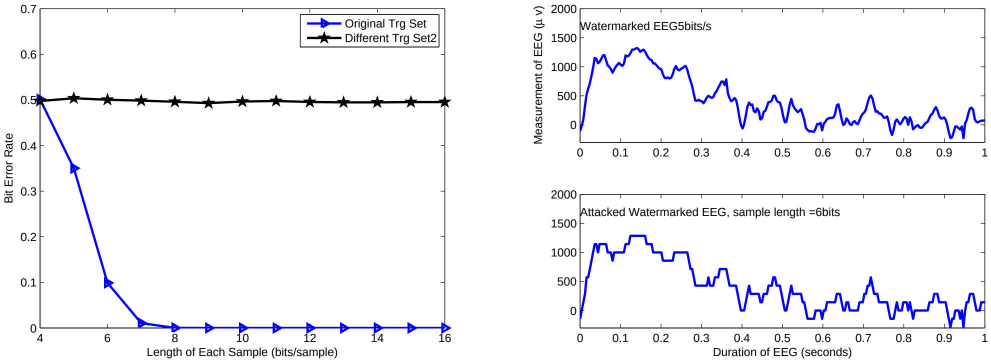


FIGURE 3: The Maximum Mutual Information between the original independent sources, and the attacker's estimated sources. The solid bar refers to the mutual information between the original sources and the grey bar to the mutual information to the attacker's best estimated sources.

Figure 4(a) shows the fractional bit error rate using the estimated and the actual demixing matrices. The figure is plotted as a function of the number of bits per sample transmitted and so reveals both robustness and also sensitivity to attack. It is evident from Figure 4(a) that without precise knowledge of the demixing matrix, recovery of the personal information in the embedded message is highly improbable.



(a) Bit error rate of reconstructed message as a function of the quantisation of the transmitted samples, using both the original and the estimated demixing matrices.

(b) Visible degradation of the EEG signal as the transmitted quantisation per sample reduces to $6bs^{-1}$.

FIGURE 4: Degradation as a consequence of reduction in transmitted bits per sample

3.3 Robustness

Medical signals are not likely to be subject to the same type of malicious attack as, say, downloaded music or video files. However benign attacks such as pre-signal processing, or downsampling of large data files to allow more efficient data transmission could be an issue. We demonstrate robustness to just one example attack: the reduction of sample resolution. The previous figure (Fig 4(a) reveals robustness of the reconstructed message as the transmitted (watermarked) biopatterns are reduced in sample dynamic range from 16 bits down to a few bits per sample. The message begins to become significantly degraded around 6 bits. However at this reolution there is a significant degradation in the EEG signal itself, as Figure 4(b) shows.

4 Conclusions

Information theory through the use of blind ICA approaches indicates a pragmatic approach to digital watermarking for privacy protection in the electronic Patient Healthcare Record. The use of ICA for steganography has the dual benefits of maximising embedding data rates, and has an inbuilt sensitivity to data snooping. We have illustrated one small but difficult example: that of embedding in single channel biomedical time series, where redundancy is low and domain-fidelity requirements are high. We have illustrated sensitivity and robustness to reasonable attack. Without a pragmatic approach to information hiding and privacy protection, the widespread deployment of the EPHR is likely to be compromised due to lack of public acceptance.

ACKNOWLEDGEMENT

This work was supported by the EU BIOPATTERN Network of Excellence, under contract 508803.

References

Bounkong, S., Saad, D. & Lowe, D. (2002). Independent component analysis for domain independent watermarking, *Lecture Notes in Computer Science 2415*, pp. 510–515.

Bounkong, S., Toch, B., Saad D. & Lowe, D. (2003). ICA for Watermarking Digital Images, *Journal of Machine Learning Research: Special Issue on Independent Component Analysis* **4**: 1471–1498.

Chen, B. & Wornell, G. (2001). Quantization index modulation: a class of provably good methods for digital watermarking and information embedding, *IEEE Transaction on Information Theory* **47**(4): 1423–1443.

Cox, I., Miller, M. & Bloom, J. (2002). *Digital Watermarking*, Morgan Kaufmann Publishers.

Moulin, P. & O'Sullivan, J. (2001). Information-theoretic analysis of information hiding. `http://www.ifp.uiuc.edu/~moulin/`, preprint Sept. 1999, revised.

Toch, B. & Lowe, D. (2005). Watermarking of medical signals, *Proceedings of the 2nd International Conference on Computational Intelligence in Medicine and Healthcare (CIMED2005)*, Lisbon, Portugal, pp. 231–236.