

XÓA WATERMARK BẰNG PHƯƠNG PHÁP BIẾN PHÂN TOÀN PHẦN KẾT HỢP PHÂN ĐOẠN WAVELET

ThS. Trương Hoàng Vinh¹

TÓM TẮT

Vào năm 1990, Tanaka đã xuất bản bài báo [8] mô tả kỹ thuật để che dấu một đoạn văn bản vào trong một bức ảnh. Công trình này đánh dấu sự ra đời của một chuyên ngành mới của khoa học máy tính – đó là watermarking. Ngày nay, chúng ta đã có một định nghĩa chính xác về watermarking, đó là sự che dấu thông tin bên trong dữ liệu đa phương tiện (hình ảnh, âm thanh, phim ảnh, ...) Một hệ thống watermarking trong xử lý ảnh phải đảm bảo cả 3 yếu tố sau: độ chắc chắn, độ bảo mật và chất lượng của ảnh. Cho đến nay, thuật toán tốt nhất cho hệ thống watermarking hình ảnh là thuật toán Broken Arrow (BA) [7] với phương pháp chen tại các bit 0. Bài báo này sẽ mô tả một kỹ thuật tấn công vào sự chắc chắn của thuật toán BA bằng việc kết hợp phương pháp biến phân toàn phần và phân đoạn wavelet.

ABSTRACT

In 1990, Tanaka has been publishing a first paper [8] to describe a method for hiding a message in a picture. This work marked the beginning a new branch of computer science. It's watermarking. Nowadays, we have an exact definition of watermarking is the art to hide information in a multimedia (image, sound, video...). A system watermarking must have three main requirements such as: robustness, security and quality. Till now, the best algorithm to watermark at zero-bit is Broken Arrow (BA) algorithms [7]. In this paper, we would like to show an attack on robustness of BA system. Our technique is the combination of the total variation and wavelet decomposition.

Keywords: watermarking, attack on robustness, total variation

I. Giới thiệu

Ngày nay cùng với sự bùng nổ của kỷ nguyên Internet, vấn đề bảo mật dữ liệu và an toàn thông tin trở nên thật cần thiết. Đối với dữ liệu đa phương tiện (như video, âm thanh, hình ảnh...) thì watermarking là một giải pháp phù hợp để bảo mật thông tin. Watermarking được định nghĩa là sự thay đổi nội dung dữ liệu đa phương tiện bằng cách chen vào một thông điệp hay một lượng dữ liệu mà mắt thường không nhìn thấy được sự thay đổi này. Từ năm 2005, công trình của Cayre và các cộng sự [6] đã nêu ra các vấn đề về an toàn bảo mật thông tin trong xử lý ảnh. Một trong các vấn đề đó là phân loại các cuộc tấn công

vào hệ thống watermark như (a) tấn công bằng xử lý ảnh (robustness attack) làm cho hệ thống không còn phát hiện ra nội dung đã chen vào dữ liệu ảnh, (b) tấn công vào hệ thống an ninh (security attack).

Thuật toán Broken Arrows (BA) [7] được sử dụng và giới thiệu trong cuộc thi BOWS-II [6] vào năm 2008. Đây được đánh giá là thuật toán tốt nhất hiện nay để bảo mật thông tin trong xử lý ảnh. Năm 2009, bài báo của Wesfeld [2] lần đầu tiên đưa ra kỹ thuật tấn công BA bằng phương pháp hồi quy tuyến tính và rò rỉ thông tin. Phần II của bài báo sẽ mô tả chi tiết thuật toán Broken Arrows (BA) và phương pháp đo lường sai số khi ảnh bị xóa watermark.

¹Giảng viên Khoa Công Nghệ Thông Tin – Đại học Mở TP HCM, vinh.th@ou.edu.vn

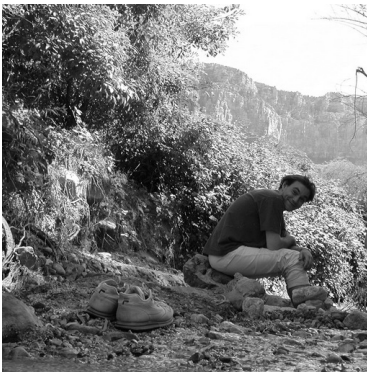
Phần III sẽ giới thiệu phương pháp biến phân toàn phần và ứng dụng vào để xóa watermark. Phần IV sẽ trình bày phương pháp xóa watermark bằng sự kết hợp giữa biến phân toàn phần và phân đoạn wavelet. Phần V sẽ trình bày kỹ thuật nâng cao chất lượng ảnh đã bị xóa watermark. Cuối cùng sẽ trình bày kết luận và triển vọng của công trình trong tương lai.

II. Thuật toán Broken Arrows (BA) và chỉ số đo lường PSNR

II.1. Thuật toán Broken Arrows

Cuộc thi BOWS-II [4] được tổ chức vào năm 2008 nhằm thách thức các nhà nghiên cứu trong lĩnh vực xử lý tín hiệu. Ban tổ chức đã đưa ra 3 bức ảnh (Hình 2.1, 2.3 và 2.4) và ẩn chứa bên trong 3 bức ảnh này là một câu thông điệp “From libit” mà mắt thường không nhìn thấy được. Nhiệm vụ của người tham gia cuộc thi là phải xóa đi hoặc làm cho hệ thống không nhận được thông điệp trên trong từng bức ảnh và phải duy trì chất lượng ảnh tối thiểu ở mức PSNR tối thiểu là 20 dB. (Xem phần II.2 – chỉ số đo lường PSNR)

Hình. 2.1: Người đàn ông



Hình. 2.2: Con cừu



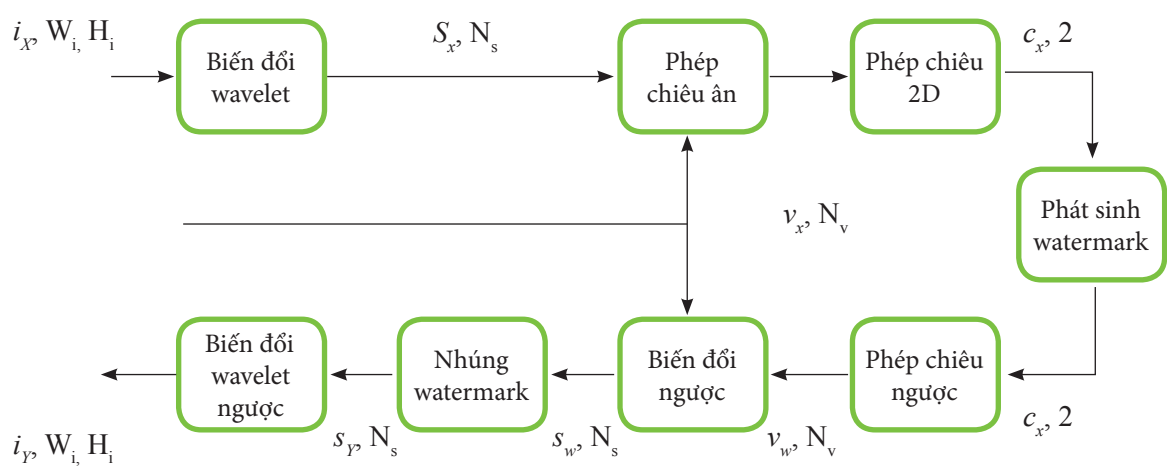
Hình. 2.3: Con suối



Quá trình nhúng và phát hiện watermark của thuật toán BA được miêu tả như Hình 2.4. Quá trình này trải qua 4 giai đoạn : giai đoạn của pixels, giai đoạn wavelet, giai đoạn tương quan và giai đoạn “Miller, Cox & Bollom” (MCB) [12]. Ba ký tự X, Y, Z được dùng theo thứ tự để biểu diễn nội dung gốc, nội dung đã watermark, và nội dung cần nhúng vào. Các thuật ngữ và ký hiệu sau đây được dùng để biểu diễn trong các giai đoạn khác nhau :

- N_s : số lượng hệ số wavelet,
- N_v : đối cực nhị phân với kích thước N_s
- “ảnh” trong miền pixels với kích thước $H_i \times W_i$
- “tín hiệu” trong miền wavelet,
- $R^{N_x}: s_y, s_x, s_w$
- “vector” trong miền tương quan
- $R^{N_v}: v_y, v_x, v_w$
- “tọa độ” trong miền mặt phẳng
- MCB, $R^2: c_y, c_x, c_w$

Hình 2.4 : Sơ đồ tóm tắt thuật toán Broken Arrow.



Ảnh được dùng trong BOWS-II là ảnh trắng đen có kích thước $H_i \times W_i$. Ta có giá trị tín hiệu watermark là sự sai biệt giữa giá trị ảnh gốc và ảnh đã watermark: $i_w = i_x - i_y$. Sự thay đổi về chất lượng ảnh giữa i_x và i_y được đánh giá bằng chỉ số PSNR (Xem II.2)

Từ ảnh gốc i_x , ta thực hiện biến đổi wavelet ở 3 mức để thu được 9 subbands HH1, HL1, LH1, HL2, HH2, LH3, HH3, LH3 và LL3. Ta chọn tất cả các hệ số wavelet của các subbands chỉ trừ của LL3. Ta có $N_s = H_i W_i (1 - \frac{1}{16})$ hệ số wavelet, các hệ số này được lưu trữ dưới dạng vector dạng cột s_x . Tín hiệu này lưu trữ trong R^{N_s} miền wavelet. Các hệ số wavelet còn lại của LL3 sẽ được lưu lại trong bộ nhớ và dùng để thực hiện bước đảo ngược để thu về ảnh đã nhúng watermark.

Tín hiệu s_x thu được từ quá trình chuyển đổi nội dung cần nhúng, và được thêm vào vector s_x để thu được tín hiệu đã watermark: $s_y = s_x + s_w$. Quá trình kết thúc khi ta ghép cặp tín hiệu s_y và hệ số wavelet của subband

II.2 Đo lường chất lượng ảnh bằng PSNR

Ảnh gốc sau khi chèn thông tin vào sẽ bị thay đổi về chất lượng ảnh, cũng như vậy, việc tấn công để xóa watermark cũng sẽ làm thay đổi chất lượng của ảnh. Có

nhiều phương pháp và chỉ số để đo lường sự thay đổi chất lượng của ảnh như PSNR, SNR,...[5] Trong phạm vi bài báo này chúng tôi chọn chỉ số PSNR cho phù hợp với tiêu chuẩn của BOWS-II [2]

Cho ảnh I có kích thước $M \times N$. $I(i,j)$ và $I'(i,j)$ lần lượt là các thành phần của ảnh gốc và ảnh sau khi nén.Sai số bình phương trung bình (MSE)

$$MSE = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \|I(i,j) - I'(i,j)\|^2 \tag{1}$$

$$PSNR = 10 \log_{10} \frac{d^2}{MSE} \tag{2}$$

Chỉ số PSNR được đo lường bằng đơn vị dB. Một ảnh có chất lượng tốt khi giá trị PSNR nằm trong khoảng từ [30, 40] dB. Ba bức ảnh dùng trong BOWS-II đều có chất lượng khá tốt vì có giá trị PSNR trung bình trên 43dB.

III. Xóa watermark bằng phương pháp biến phân toàn phần

III.1 Biến phân toàn phần

Phương pháp biến phân toàn phần đo lường tổng dao động của một tín hiệu. Trong xử lý ảnh, tín hiệu phụ thuộc vào số lượng chiều dài của các dòng. Biến phân toàn phần của hàm số f được định nghĩa bởi công thức :

$$f_{VT} = \int_{-\infty}^{+\infty} |f'(t)| dt \tag{3.1}$$

Cho ảnh với kích thước $N \times N$, ta định nghĩa ∇u là một vector $Y = X \times X$ được xác định bởi :

$$(\nabla u)_{i,j} = \left((\nabla u)^1_{i,j}, (\nabla u)^2_{i,j} \right), \text{ với}$$

$$(\nabla u)^1_{i,j} = f(x) = \begin{cases} u_{i+1,j} - u_{i,j}, & \text{nếu } i < N \\ 0, & \text{nếu } i = N \end{cases}$$

$$(\nabla u)^2_{i,j} = f(x) = \begin{cases} u_{i,j+1} - u_{i,j}, & \text{nếu } j < N \\ 0, & \text{nếu } j = N \end{cases}$$

$$(3.2)$$

Với $i, j = 1, \dots, N$

Hàm biến phân toàn phần của u xác định bởi

$$J(u) = \sum_{1 \leq i, j \leq N} |(\nabla u)_{i,j}| \quad (3.3)$$

Ta có chuẩn $\|y\| = \sqrt{y_1^2 + y_2^2}$, cho mỗi cặp $y = (y_1, y_2) \in \mathbf{R}^2$

Ta giới thiệu hàm phân kỳ rời rạc $\text{div}: Y \rightarrow X$, và được xác định bằng:

$$(\text{div } p)_i(i, j) = \begin{cases} (p_1(i, j))^T 1 - p_1(i-1, j)^T 1 & \text{nếu } 1 < i < N \\ (p_1(i, j))^T 1 & \text{nếu } i = 1 \\ (p_1(i, j))^T 1 - p_1(i, j-1)^T 1 & \text{nếu } i = N \end{cases}$$

$$+ \begin{cases} (p_2(i, j))^T 2 - p_2(i, j-1)^T 2 & \text{nếu } 1 < j < N \\ (p_2(i, j))^T 2 & \text{nếu } j = 1 \end{cases}$$

$$(3.4)$$

Với mỗi $p = (p^1, p^2) \in Y$

III.2 Xóa watermark bằng thuật toán biến phân toàn phần

toàn phần trong xử lý ảnh nhằm để khử nhiễu là ROF (Rudin, Osher, and Fatemi) [3], với ràng buộc làm giảm mức tối thiểu:

$$\min_{u \in X} \left(\int_{\Omega} |\nabla u| + \lambda \int_{\Omega} (u - g)^2 dx \right) \quad (3.5)$$

Tương đương với :

$$u = \arg \min_{u \in X} \left(\|u - g\|_2^2 - N \sigma^2 \right) \quad (3.6)$$

Với σ biểu diễn kích thước giá trị sai số ε của dữ liệu trong công thức (3.6), $\|\cdot\|$ biểu diễn chuẩn Euclidean trong $\mathbf{R}^2$, và $\|\cdot\|_X$ biểu diễn chuẩn trong X . $\lambda > 0$ là tham số do người dùng lựa chọn để điều khiển lượng thông tin nhiễu cần xóa đi. Giá trị

tham số λ tỷ lệ nghịch với lượng thông tin xóa đi trên ảnh.

Dựa trên ý tưởng của ROF, Chambolle [1] đã đề xuất ra mô hình lọc nhiễu cho ảnh bằng cách giải phương trình (3.7):

$$u = g - \lambda \text{divp } n^2$$

(3.7)

Ta có $dt = \frac{1}{4}, p^0 = 0$, với bất kỳ $n \geq 0$,

$$p_{i,j}^{n+1} = \frac{p_{i,j}^n + dt \left(\nabla \left(\text{divp}^n - \frac{g}{\lambda} \right) \right)_{ij}}{\left(\left[p_{i,j}^n + dt \left| \nabla \left(\text{divp}^n - \frac{g}{\lambda} \right) \right|_{i,j} \right]_{i,j} \right)}$$

(3.8)

Thuật toán tính giá trị ảnh bị nhiễu u bằng cách xác định giá trị p^n , điều kiện để thoát khỏi vòng lặp khi giá trị giá trị lớn nhất của $p_{i,j}^n$ và $p_{i,j}^{n+1}$ phải nhỏ hơn $\frac{1}{100}$.

Ta xem dữ liệu đã nhúng vào ảnh bằng thuật toán BA như là một giá trị nhiễu. Ta tiến hành xóa watermark bằng

cách chọn giá trị λ và so sánh kết quả với hệ thống BA cho đến khi nào không còn phát hiện ra dữ liệu đã watermark. Ta có kết quả như bảng 3.1 :

IV Kết hợp phân đoạn wavelet và biến phân toàn phần

IV.1 Tấn công bằng phân đoạn wavelet.

Bảng 3.1 : Kết quả xóa watermark bằng phương pháp biến phân toàn phần

Ảnh	Giá trị λ	PSNR
Con suối	0.11	19.97 dB
Con cừu	4.9	20.76 dB
Người đàn ông	3.7	20.24 dB
Giá trị trung bình		20.32 dB

Thuật toán BA sử dụng các hệ số wavelet (phân đoạn ở mức 3) để nhúng dữ liệu vào ảnh. Một cách hiển nhiên khi ta vô hiệu hóa hệ số này bằng cách áp giá trị 0 cho từng hệ số thì hệ thống sẽ không thể nhận dạng được dữ liệu đã nhúng. Tuy nhiên lúc này chỉ số PSNR sẽ không đảm

bảo yêu cầu là lớn hơn 20dB vì ảnh đã bị thay đổi đáng kể về mặt chất lượng. Ở đây chúng tôi chỉ tiến hành phân đoạn wavelet ở mức 2 và mức 3 nhằm duy trì chỉ số PSNR ở mức cao. Kết quả tấn công bằng phân đoạn wavelet sau khi đã kiểm tra với hệ thống BA cho ta kết quả như bảng 4.1:

Bảng 4.1 Kết quả xóa watermark bằng phương pháp phân đoạn wavelet

	Mức 2	Mức 3
Con suối	21.42 dB	19.08 dB
Con cừu	22.68 dB	21.31 dB
Người đàn ông	21.32 dB	20.04 dB
Giá trị trung bình 21.80 dB		20.14 dB
Kết quả xóa watermark	Không thành công	Thành công

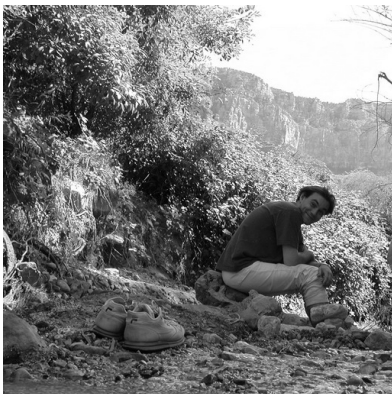
IV.2 Kết hợp phân đoạn wavelet và biến phân toàn phần

Sau khi tiến hành tấn công riêng lẻ thuật toán BA bằng phương pháp biến phân toàn phần và phân đoạn wavelet. Chúng tôi kết hợp cả hai phương pháp trên nhằm đảm bảo chất lượng ảnh và sự thành công của cuộc tấn công. Ảnh sẽ lọc nhiễu bằng chỉ số λ . Ở đây ta chọn λ tương đối lớn nhằm mục đích chỉ xóa một lượng nhỏ thông tin. Phương pháp này có ý nghĩa là ta chỉ lọc nhiễu một phần nhỏ lượng thông tin của ảnh. Sau đó sẽ tiến hành phân đoạn wavelet ở mức 2 và đặt tất cả các hệ số wavelet này bằng giá trị 0. Kết quả thu được sau khi kết hợp hai phương pháp biến phân toàn phần và phân đoạn wavelet như bảng 4.2 và Hình 4.1

Bảng 4.2: Kết quả xóa watermark bằng phương pháp biến phân toàn phần và phân đoạn wavelet

Ảnh	Giá trị λ	PSNR
Con suối	17.99	20.77 dB
Con cừu	41.5	21.49 dB
Người đàn ông	23.58	20.39 dB
Giá trị trung bình		20.88 dB

Hình 4.1 : Ba bức ảnh của BOWS-II sau khi đã xóa watermark bằng phương pháp biến phân toàn phần kết hợp phân đoạn wavelet.



VI. Kết luận – Thảo luận

Chúng tôi vừa trình bày kỹ thuật tấn công thuật toán BA bằng phương pháp biến phân toàn phần kết hợp phân đoạn wavelet. Thuật toán BA được đánh giá có độ bảo mật tốt nhất hiện nay dành cho zero-bit. Độ bảo mật của thuật toán đã được kiểm tra qua cuộc thi BOWS-II. Vào cuối năm 2009, Westfeld [9] đã công bố kỹ thuật tấn công BA bằng phương pháp hồi quy tuyến tính. Triển vọng trong tương lai có thể nghiên cứu thay thế biến đổi wavelet của BA bằng một số miền biến đổi khác như Contourlet và Ridgelet [9, 10, 11]. Và cũng tương tự như vậy, ta phải đánh giá độ bảo mật bằng thông qua các kỹ thuật tấn công như đã trình bày.

TÀI LIỆU THAM KHẢO

1. A. Chambolle, "An Algorithm for Total variation Minimization and Applications", Journal of Mathematical Imaging and Vision 20 : 89-97, 2004.
2. A. Westfeld, "A Regression-based Restoration Techniques for Automated Watermark

- Removal”, TU Dresden, Multimedia and Security ACM Workshop MM-SEC2008, Oxford, United Kingdom, 22-23, September 2008.
3. Rudin LI, Osher S, Fatemi E, “Nonlinear total variation based noise removal algorithms”, *Physica D*. 1992;60(1-4) :259-268.
 4. BOWS-2 : The second Break Our Watermarking System Contest, 17/07/2007-17/04/2008, Organised within the activity of the Watermarking Virtual Laboratory (Wavila) of the European Network of Excellent ECRYPT - <http://bows2.gipsalab.inpg.fr>.
 5. E.Marini, F.Autrusseau, P.L. Callet, P.Campisi, “Evaluation of standard watermarking techniques”, in *Proceedings of SPIE*, Vol 6505, pp. 65050O, 2007.
 6. Cayre, C. Fontaine and T. Furon “Watermarking Security, F. Digital Audio Watermarking Techniques and Technologies: Applications and Benchmarking, Idea Group Publishing, (2007)
 7. T. Furon and P. Bas, “Broken Arrows”, *EURASIP Journal on Information Security*. Volume 2008. Article No. 6. ISSN :1687-4161.
 8. K. Tanaka, Y. Nakamura, and K. Matsui, “Embedding Secret Information into a Dithered Multi-level Image”, In *Proceedings of 1990*.
 9. E.J. Candès and D.L. Donoho, “Ridgelets : the key to high dimensional intermittency”, *Philosophical Transactions of the Royal Society of London A*, 357 :2495-2509, 1999.
 10. E. J. Candès and D. L. Donoho, “Curvelets - a surprisingly effective nonadaptive representation for objects with edges”, In A. Cohen, C. Rabut, and L.L. Schumaker, editors, *Curve and Surface Fitting : Saint-Malo 1999*, Nashville, TN. Vanderbilt University Press , 1999.
 11. M. N. Do and M. Vetterli. “Contourlets”, In J. Stoeckler and G. V. Welland, editors, *Beyond Wavelets*. Academic Press, 2003.
 12. M. Miller and J. Bloom, “Computing the probability of false watermark detection,” in *Proceedings of the 3rd International Workshop on Information Hiding (IH ‘99)*, A. Pfitzmann, Ed., vol. 1768 of *Lecture Notes in Computer Science*, pp. 146–158, Springer, Dresden, Germany, September 1999.