Hindawi Publishing Corporation
EURASIP Journal on Information Security
Volume 2007, Article ID 98684, 8 pages
doi:10.1155/2007/98684

Research Article

Design and Analysis of the First BOWS Contest

A. Piva[1] and M. Barni[2]

[1] *Department of Electronics and Telecommunications, University of Florence, 50139 Florence, Italy*
[2] *Department of Information Engineering, University of Siena, 53100 Siena, Italy*

Correspondence should be addressed to A. Piva, piva@lci.det.unifi.it

Received 20 June 2007; Accepted 29 August 2007

Recommended by F. Pérez-González

The break our watermarking system (BOWS) contest was launched in the framework of the activities carried out by the European Network of Excellence for Cryptology ECRYPT. The aim of the contest was to investigate how and when an image watermarking system can be broken while preserving the highest possible quality of the content, in the case the watermarking system is subject to a massive worldwide attack. The great number of participants and the echo that the contest has had in the watermarking community contributed to make BOWS a great success. From a scientific point of view, many insights into the problems attackers have to face with when operating in a practical scenario have been obtained, confirming the threat posed by the sensitivity attack, which turned out to be the most successful attack. At the same time, several interesting modifications of such an attack have been proposed to make it work in a real scenario under limited communication and time resources. This paper describes how the contest has been designed and analyzes the general progress of the attacks during the contest.

1. INTRODUCTION

The first break our watermarking system (BOWS) contest has been organized by the Watermarking Virtual Laboratory (WAVILA) of the European Network of Excellence ECRYPT [1]. ECRYPT is a network of excellence funded within the Information Society Technologies (IST) Programme of the European Commission's Sixth Framework Programme (FP6), launched on February 2004 and lasting on July 2008. Its objective is to intensify the collaboration of European researchers in information security, and more in particular in cryptology and digital watermarking. The activities of the ECRYPT network of excellence are organized into five virtual laboratories, the first four on cryptographic research activities, and the last one on watermarking and perceptual hashing (WAVILA). WAVILA aims at building tools and techniques for assessing the security aspects of watermarking and perceptual hashing, to design advanced algorithms with a well-defined security level, to design protocols, both standalone as well as integrated in cryptographic protocols, and to develop methods and techniques for efficient and secure implementations.

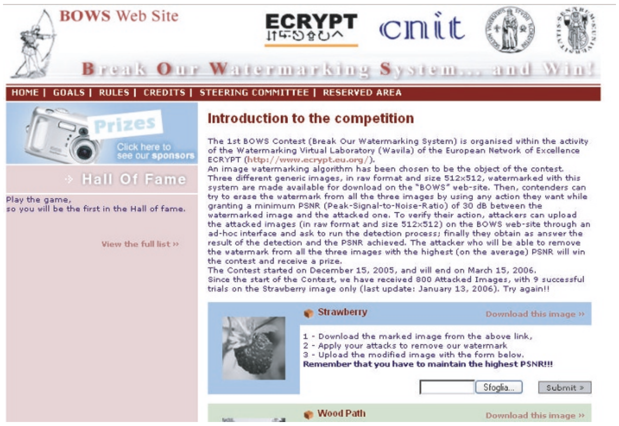
In the framework of WAVILA activities, it was proposed to launch a contest that was named *break our watermarking system (BOWS)*. As suggested by the name, BOWS was designed to allow the researchers interested in watermarking to investigate how and when an image watermarking system can be broken while preserving the highest possible quality of the modified content, in case that the watermarking system is subjected to a worldwide massive attack. The BOWS contest was not intended to prove how well-performing a watermarking system is, but it was expected by means of this action to better understand which are the disparate possible attacks, perhaps unknown at the moment of the start of the contest, the BOWS participants could carry out to perform their action and comprehend the degree of difficulty of breaking the embedded watermark.

In addition, the contest was designed to study if and how much the knowledge of the watermarking algorithm is useful for watermark removal. In fact, according to an approach similar to the Kerckhoffs principle [2] adopted in cryptography (stating that the security of a cryptographic scheme should not rely on the secrecy of the cryptographic algorithm but on one or more secret keys), watermarking security is often analyzed by assuming that the attacker can have full knowledge of the watermarking scheme and then he/she can explicitly exploit such a knowledge to design a proper attacking strategy. This assumption is based on the concept that the knowledge of the details of the watermarking algorithm helps a lot the attacker. To study the importance of the knowledge of the watermarking algorithm for watermark removal, it was decided to divide the contest in two phases: in the first phase the watermarking algorithm was not revealed in contrast to the Kerckhoffs principle, in the second phase the algorithm was made public, to allow the researchers to sharpen their attacks with more information about the watermarking scheme.

This paper describes how the contest has been designed and analyzes the general progress of the attacks during the two phases composing the contest itself.

2. DESIGN OF THE CONTEST

The general form of the contest was conceived in the following way: three grayscale images were watermarked with a one-bit watermarking algorithm. The watermarked images were available for download on the BOWS website at the address http://lci.det.unifi.it/BOWS, whose homepage is shown in Figure 1. After downloading, contenders were allowed to try to erase the embedded watermark from the three images by using any action they wanted while granting a minimum PSNR of 30 dB between the watermarked image and the attacked one. Note that the adoption of the PSNR as quality measure automatically excluded from the set of available attacks the geometrical modifications, since even a small geometrical distortion, like a shift by one pixel, heavily affects the PSNR. To verify their action, attackers were asked to upload each of the three images (still in raw format and size 512 × 512) on the BOWS website through an ad-hoc interface shown in Figure 2(a) to ask to run the detection process; finally they obtained as answer the result of the detection and the PSNR achieved (Figure 3). In case of successful attack, the thumbnail of the attacked image exhibited the stamp "Passed," as in Figure 2(b). When a BOWS participant succeeded to remove the watermark from all the three images, he/she was asked to register in the hall of fame. The best performances in terms of PSNR (average PSNR on the three images when watermark deletion has been successful over all of them) were stored by the system to fill in a rank list updated in realtime. The attacker able to remove the watermark from all the three images with the highest average PSNR was the winner of the contest. At the beginning of the contest, a limit of 30 uploads/day was fixed. To check it and to log the working out of the contest, all the uploads were recorded, according to the IP address of the client connecting to the BOWS server. Afterwards this limit was removed and set to 5000 images per day (per user).


BOWS Web Site
ECRYPT
cnit
Break Our Watermarking System... and Win!
HOME | GOALS | RULES | CREDITS | STEERING COMMITTEE | RESERVED AREA
Prizes
Click here to see our sponsors
Hall Of Fame
View the full list »
Download this image »
Sfoglia...
Submit »
Wood Path
Download this image »

Play the game,
so you will be the first in the Hall of fame.

Introduction to the competition

The 1st BOWS Contest (Break Our Watermarking System) is organised within the activity of the Watermarking Virtual Laboratory (Wavila) of the European Network of Excellence ECRYPT (http://www.ecrypt.eu.org/).
An image watermarking algorithm has been chosen to be the object of the contest. Three different generic images, in raw format and size 512×512, watermarked with this system are made available for download on the "BOWS" web-site. Then, contenders can try to erase the watermark from all the three images by using any action they want while granting a minimum PSNR (Peak-Signal-to-Noise-Ratio) of 30 dB between the watermarked image and the attacked one. To verify their action, attackers can upload the attacked images (in raw format and size 512×512) on the BOWS web-site through an ad-hoc interface and ask to run the detection process; finally they obtain as answer the result of the detection and the PSNR achieved. The attacker who will be able to remove the watermark from all the three images with the highest (on the average) PSNR will win the contest and receive a prize.
The Contest started on December 15, 2005, and will end on March 15, 2006.
Since the start of the Contest, we have received 800 Attacked Images, with 9 successful trials on the Strawberry image only (last update: January 13, 2006). Try again!!

Strawberry

1 - Download the marked image from the above link,
2 - Apply your attacks to remove our watermark
3 - Upload the modified image with the form below.
Remember that you have to maintain the highest PSNR!!!

Figure 1: The home page of the BOWS website, available at http://lci.det.unifi.it/BOWS.

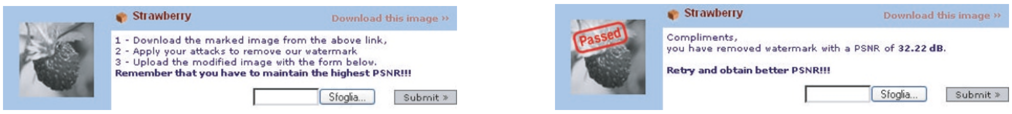
2.1. Choice of the watermarking algorithm

The choice of the watermarking algorithm was dictated by the following factors: (i) desire to test a *modern* system based on the theory of side informed watermarking; (ii) necessity of obtaining the consensus of the inventors of the watermarking systems to use it in the contest; (iii) necessity of using a complete system, including the exploitation of the human visual system for better watermark hiding. With the above ideas in mind, the chosen watermarking algorithm was the one designed by Miller et al. [3], with the agreement of the authors, to be the object in the contest. While a detailed description of the algorithm can be found in the original paper by Miller et al. [3], here it may be interesting to recall that the watermark is embedded in the block DCT domain, in the low-range portion of the spectrum (specifically, by considering a zigzag scanning of the block DCT coefficients, the watermark was inserted in the coefficients ranging from the second to the thirteen). Before embedding the DCT, coefficients were scrambled to avoid that bits were associated to particular image area, thus weakening some of them and reinforcing others.

The watermarking strength was adjusted according to Watson's model [4], for a final PSNR ranging from 42 to 46 dBs (see Table 1). Whereas the original algorithm was conceived as a multibit system, we turned it into a one-bit scheme by inserting within the image a particular codeword and asking to the detector to check whether the extracted content was equal to the original one. No redundancy was introduced in this step, since we decided to apply all the protection to the individual bits, by means of a dirty-paper trellis mechanism (see the original paper for more details). While this choice simplified the analysis of the false positive detection rate and the transformation of the original multibit scheme into a one-bit algorithm, the resulting algorithm resulted more vulnerable given that even changing a few coefficients was enough to inhibit the correct watermark detection.

2.2. Choice of the host images

Three grayscale images, with different visual characteristics (*Strawberry*, *Wood Path*, and *Church* shown in Figure 4) in raw format and size 512 × 512, were chosen for watermarking. The three images were selected so to represent three different classes of images, namely images characterized by low activity (the *Strawberry*), images with strong regular structures (the *Church*), and images with irregular textured content (the *Wood Path*). As it was evident from the contest results (and as it was expected), the *Strawberry* image was the



Strawberry

Download this image »

1 - Download the marked image from the above link,
2 - Apply your attacks to remove our watermark
3 - Upload the modified image with the form below.
Remember that you have to maintain the highest PSNR!!!

(a)

Strawberry

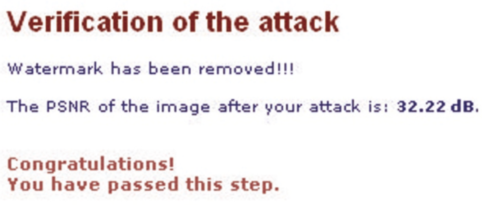
Download this image »

Compliments,
you have removed watermark with a PSNR of **32.22 dB**.

Retry and obtain better PSNR!!!

(b)

Figure 2: The BOWS interface to upload the attacked images and run the detector, as it appears before a successful attack (a), and after a successful attack (b) on the image *Strawberry*. Note that in case of successful attack the thumbnail of the image shows the stamp "Passed," and that the PSNR value achieved by the current attack is recorded.



Verification of the attack

Watermark has been removed!!!

The PSNR of the image after your attack is: **32.22 dB**.

Congratulations!
You have passed this step.

Figure 3: The answer of the detector, with the PSNR value achieved by the current attack.

Table 1: PSNR values of the three watermarked images with respect to the original ones.

Image	*Strawberry*	*Wood Path*	*Church*
PSNR	42.144 dB	46.198 dB	44.382 dB

easiest to attack by means of standard image processing tools, while no particular difference could be observed with regard to sensitivity attacks. The three original images were watermarked with the adopted system obtaining the watermarked versions exhibiting a peak signal-to-noise ratio (PSNR) with respect to the original ones included between 42 and 46 dB, as it is described in Table 1. As it is shown, the distortions introduced by the watermark embedding are lower for the *Wood Path* image, whereas they are higher for the *Strawberry* image.

2.3. *False positives*

An interesting question regards the false positive rate. The detection parameters were set by fixing a (theoretical) false positive rate of 2^{-40}, however such a rate is computed by considering nonattacked images, whereas one may argue that pirates may also be interested in generating falsely watermarked images. For this reason at the beginning it was planned to create a section of the contest devoted to the generation of false positive images, however for the sake of simplicity this idea was abandoned. Nevertheless, in order to keep the false positive rate under examination, we recorded all the images for which the detector gave a positive answer with a PSNR lower than 10 dB; at the end of the contest, about 200 images still watermarked even with a PSNR lower than 10 dB were recorded; this means that the rate of 2^{-12} was obtained; however, we can observe that these images do not allow to estimate the true false positive rate, since most of them are watermarked images that retained the watermark even in the presence of a very strong attack (see [5] for an interesting analysis of this aspect).

3. ANALYSIS OF THE CONTEST

As it has already been described, the contest consisted of two phases: in the first phase the watermarking algorithm was secret, whereas in the second phase it was made public. The official winner of the contest prize was decided to be the winner of the first phase. In the following the two Phases are analyzed, and a comparison between them is carried out.

3.1. *First phase of the contest*

The first phase of the BOWS contest started on December 15, 2005, and ended on March 16, 2006. At the beginning of the contest the participants were able to remove the watermark only on the image *Strawberry*: it seems that in this image it was easier than in the other images to find and modify the watermarked features. It was then decided to remove the limit on the maximum number of attacks per day in order to allow the attackers to carry out also sensitivity attacks [6–8] (actually, the limit was not removed, but fixed to a value equal to 5000 attacks/day). Thanks to this modification and to the growing advertisement of the contest, the number of participants and uploaded attacked images increased alot. At the end of the first phase of the contest, 72074 attacked images were uploaded from more than 300 IP addresses; in 10034 of them (corresponding to the 13.9% of all the received images) the watermark was erased while granting a minimum PSNR of 30 dB between the watermarked image and the attacked one. However, only 10 participants succeeded to remove the watermark from all the 3 watermarked images, and registered their data in the hall of fame. The steering committee responsible to rule the BOWS contest, according to the recorded results, confirmed that the winner was the team held by Scott Craver, from Binghamton University, with the following results: PSNR of the image *Strawberry* = 39.67 dB, PSNR of the image *Wood Path* = 39.65 dB, and the PSNR of the image *Church* = 38.45 dB.

By analyzing the hall of fame at the end of the first phase, it is possible to note that most of the successful attacks have been registered in the last three or four days of the contest; seven attacks obtained an average PSNR lower than 31 dB, and only three were able to exceed 36 dB. The complete hall of fame, as it appeared at the end of the first phase of the contest, is given in Table 2.

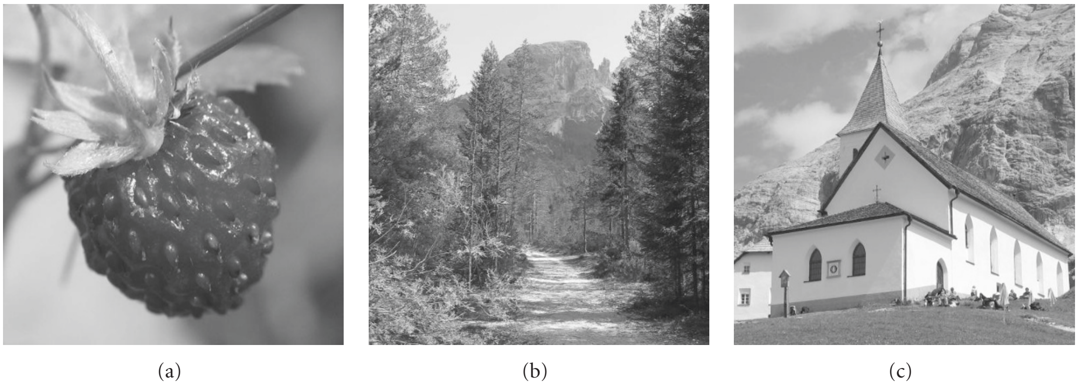


Figure 4: The three original images, *Strawberry*, *Wood Path*, and *Church*, used for the contest.

Table 2: Hall of fame at the end of the first phase of the BOWS contest.

Date	Participant	Institute	Av PSNR (dB)
03/14/06	Team Craver	Binghamton University	39.22
03/14/06	Team Craver	Binghamton University	37.94
03/16/06	J. Earl	Cambridge University	37.27
03/14/06	G. Le Guelvouit	Capgemini	35.24
03/13/06	M. Noisternig	Uni Salzburg	34.57
03/15/06	J. Earl	Cambridge University	34.49
03/13/06	G. Le Guelvouit	Capgemini	33.85
03/15/06	Team Dugelay	Eurecom Image	33.65
03/15/06	P. J. Doets	Delft University of Technology	33.54
02/12/06	S. Craver	Binghamton University	33.23
03/15/06	A. Westfeld	TU Dresden	32.86
03/14/06	A. Westfeld	TU Dresden	32.45
03/13/06	P. J. Doets	Delft University of Technology	31.60
03/15/06	J. Earl	Cambridge University	30.73
03/08/06	D. Bogumil	Warsaw University of Technology	30.61
03/16/06	I. Ocnarescu-A. David	University Politehnica of Bucharest	30.59
03/13/06	F. Cayre	LIS/INP Grenoble	30.59
02/02/06	G. Le Guelvouit	Capgemini	30.48
03/16/06	I. Ocnarescu-A. David	University Politehnica of Bucharest	30.44
02/02/06	A. Westfeld	TU Dresden	30.40

3.2. Second phase of the contest

After the three months of the contest, it was revealed that the watermarking algorithm used to embed the watermark into the three images was the one developed by Miller et al. [3]. Then, the BOWS website remained open for other three months for the second phase of the contest during which the researchers were allowed to sharpen their attacks by exploiting the knowledge about the adopted watermarking scheme. The hall of fame was not erased, but the participants entered in the rank in the second phase of the contest were highlighted by a different notation in the list. During these further three months, the BOWS server received 721734 attacked images from more than 100 IP addresses; in 20666 of them (corresponding to the 2.9% of all the received images) the watermark was removed while granting a minimum PSNR of 30 dB. In this second phase, 16 participants succeeded to remove the watermark from all the 3 watermarked images, and registered their data in the hall of fame (some of them succeeded several times).

The contender reaching the highest PSNR value was Andreas Westfeld, from TU Dresden, that was also the most active in the upload of attacked images, so that we were also constrained to fix a limit, even though high (3000 attacks), to the number of images uploaded by an IP address each day not to overload the server. Andreas Westfeld at the end of the contest obtained excellent values of the PSNR: for the image *Strawberry* 60.74 dB, for the image *Wood Path* 57.05 dB, and for the image *Church* 57.29 dB, with an average PSNR value on the three images of 58.07 dB. By analyzing the hall of fame concerning the second phase, it is possible to note that all the best results have been achieved by A. Westfeld: if we exclude him, the best result was the one by Michel Chekrallah (EPSIL Lebanon) that reached 36.76 dB, whereas all the other results are slightly higher than the minimum threshold of 30 dB, being included between 30.11 dB and 31.53 dB, as shown in Table 3, where the 25 records composing the hall of fame of the second phase of the contest are shown.

3.3. First phase versus second phase

This section is devoted to the comparison between the results obtained during the two phases of the contest, to try to understand if the knowledge of the watermarking scheme in the last three months has been useful for the contenders. By analyzing the results, summarized in Table 4, it is possible first of all to note that a limited number of participants succeeded to remove the watermark from all the three images, demonstrating that the adopted watermarking scheme is highly robust. In fact, at the end of the first phase of the contest the hall of fame was composed by only 20 records, whereas in the second part 25 new successful attacks entered in. However, most of the contenders registered more than once in the database, since they were able to increase the performance of their attacks, so that actually only 10 participants succeeded in the first phase, and 17 in the second one. In particular, it is interesting to note that the best results were obtained by researchers expert and well known in the watermarking area. The attacks have been carried out by a high number of clients in the first phase; in many cases, only a limited number of trials were applied by the contender, without removing the watermark, after which the contender refrained from continuing the contest. This fact can be explained by assuming that the first trials were carried out by people without experience on watermarking that perceived the contest too difficult for their skills, and thus after a few trials decided to stop their participation to the contest. In the second phase of the contest, a lower number of contenders participated, but with greater experience. The number of attacks in the second phase was ten times the attacks in the first one; however, the successful attacks were only twice as much, so that the percentage of successes decreased a lot from 13.9% to only 2.86%, showing that in the second part of the contest the sensitivity attack [6–8], based on a high number of uploads and small changes in the parameters controlling the attack, was heavily applied. This fact is confirmed when the number of images uploaded by each IP address is analyzed. As a matter of fact, the contest log files show that most images have been received by computers used by A. Westfeld; in particular, his attacks definitely prevailed in the second part of the contest (we estimated that he uploaded more than 600.000 images), whereas in the first one, his images represented about one half of the attacks. These results indicate that A. Westfeld made massive use of the sensitivity attack during the contest, with particular reference to the second phase.

In Table 5, the ten best results of the BOWS hall of fame, at the end of the six months of the contest, are shown; the dates in italic highlight the successful attacks carried out in the second phase of the contest. Within this ranking, five results belong to the first phase, and five to the second one, so that it appears that both the two phases of the contest achieved good results, even if the best three results were obtained by A. Westfeld in the second period.

4. ANALYSIS OF THE ATTACKED IMAGES AND THE MOST SUCCESSFUL ATTACKS

In this section we give some more details about the more frequent kind of attacks that have been applied during the contest, and most of all, about the quality of the attacked images.

4.1. Analysis of the attacks

Though several kinds of attacks were applied during the contest, the most successful ones were all linked to the sensitivity attack. This fact seems to confirm the threat posed by this kind of attack and the rather good maturity reached by the watermarking field with respect to conventional image processing algorithms. The above result is confirmed by the fact that until the limit of 30 attacks per day was active no attacker was able to enter the hall of fame.

From a scientific point of view, the most relevant results regarded the development of techniques that helped to speed up the sensitivity attack, given the huge computational and communication resources necessary for the implementation of such an attack in its original form.

By looking at the quality of the attacked images, we can see that often the attacks concentrated on very small areas of the image (see Figure 5, where a particular of an attacked image *Strawberry* uploaded by A. Westfeld and exhibiting a PSNR value of 41.21 dB is shown). This fact depends on the choice of detecting the presence of the watermark only if all the bits of the embedded message were correctly decoded. It is clear that with this choice the optimum strategy from a PSNR point of view consists in attacking only the blocks bearing the weakest bit, whose position could be found by some sort of sensitivity attack. More details about the most successful attacks can be found in other papers of the present issue of the EURASIP Journal on Information Security [9–12], or in the papers presented during a special session of Security, Steganography, and Watermarking of Multimedia Contents IX conference, held in January 2007 [5, 13–17].

4.2. Quality of attacked images

Concerning the evaluation of the quality of the attacked images, the first result to be highlighted is that the extended phase of the contest allowed to increase the mean PSNR of the attacked images from 39.22 dB up to 58.07 dB. Here, the increase of the PSNR during the contest representing the measured perceptual quality of the attacked images with respect to the watermarked versions is analyzed in more detail.

The quality of the attacked images is now evaluated from a chronological point of view, by taking into account the three images uploaded by the ten contenders that achieved the best average PSNR values, summarized in Table 5. It can be noted that in this ranking, 5 participants succeeded in the first phase, and 5 in the second phase; to highlight more the increase of performance achieved during the contest, we have ordered the results not according to the decreasing average

Table 3: Hall of fame of the second phase of the BOWS contest.

Date	Participant	Institute	Av PSNR (dB)
06/12/06	Andreas Westfeld	TU Dresden	58.07
04/05/06	Andreas Westfeld	TU Dresden	51.08
03/24/06	Andreas Westfeld	TU Dresden	41.00
03/23/06	Andreas Westfeld	TU Dresden	37.78
06/02/06	Michel Chekrallah	EPSIL Lebanon	36.76
04/10/06	Jonathan Vayn	Eurecom	31.53
06/15/06	R.Vigoulette-S.Francfort	France Telecom	31.12
06/14/06	R.Vigoulette-S.Francfort	France Telecom	31.05
04/09/06	Bakhtiari Ahmad-Reza	Eurecom	30.94
04/17/06	Javier Ramis	ETSET Barcelona	30.79
05/20/06	Ehab M. Ghanem	AAST	30.79
04/09/06	Cosson Romuald	Eurecom	30.69
04/09/06	Cosson Romuald	Eurecom	30.66
04/10/06	Dilmahomod Waziim	Eurecom	30.61
04/10/06	Amjad Zoghbi	Eurecom	30.53
04/10/06	KADRI Hiba	Eurecom	30.52
04/10/06	Eduardo Ramirez	Eurecom	30.50
04/10/06	Pierre	Eurecom	30.47
04/10/06	Dilmahomod Waziim	Eurecom	30.42
04/09/06	Kadri	Eurecom	30.41
04/10/06	Jaroslaw Syrokosz	Eurecom	30.40
04/10/06	Ghayati	Eurecom	30.28
04/10/06	Oscar Morales	Eurecom	30.26
06/12/06	R. Vigoulette-S. Francfort	France Telecom	30.21
04/11/06	Olivier Beauvais	Eurecom	30.11

Table 4: Summary of the results in the first phase versus in the second phase of BOWS contest.

	First phase	Second phase
IP addresses	300	100
Attacks	72074	721734
Successes	10034	20666
% successes	13.90%	2.86%
Records in hall of fame	20	25
Participants in hall of fame	10	17
Av. PSNR of rank no. 1	39.22 dB	58.07 dB

PSNR value, like in Table 5, but in chronological ascending order. In this way, the first five values represent results of the first phase of the contest, and the last five are results of the second phase. In Figure 6 a summary of these results is given: the PSNR values between the attacked images and the watermarked ones of each of the three images are represented in the graphic. It is possible to note that all the improvement of the results is due to the attacks carried out by Westfeld; as a matter of fact, the only result of the second phase not achieved by Westfeld, that is the one obtained by Chekrallah, is comparable to all the results achieved by the best participants of the first phase.

Table 5: The ten best results of the BOWS hall of fame, at the end of the two phases of the contest. The attacks carried out in the second phase of the contest are the ones with the dates in Italic.

Date	Participant	Institute	Av PSNR (dB)
06/12/06	A. Westfeld	TU Dresden	58.07
04/05/06	A. Westfeld	TU Dresden	51.08
03/24/06	A. Westfeld	TU Dresden	41.00
03/14/06	Team Craver	Binghamton Uni.	39.22
03/14/06	Team Craver	Binghamton Uni.	37.94
03/23/06	A. Westfeld	TU Dresden	37.78
03/16/06	J. Earl	Cambridge Uni.	37.27
06/02/06	M. Chekrallah	EPSIL Lebanon	36.76
03/14/06	G. Le Guelvouit	Capgemini	35.24
03/13/06	M. Noisternig	Uni Salzburg	34.57

4.2.1. Impact of distortion measure

During the contest, the choice of measuring the quality of the attacked images by means of PSNR has sometimes been criticized, since it is well known that such a measure does not reflect the way the human visual system perceives image degradation. In order to verify the correctness of the adopted measure, we rearranged the hall of fame by considering different distortion measures. Specifically the following measures

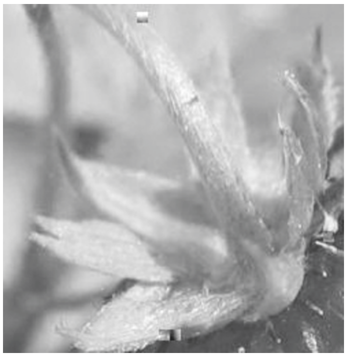


Figure 5: A particular of an attacked version of the image *Strawberry* uploaded by A. Westfeld and exhibiting a PSNR value of 41.21 dB; a modification of only a small number of blocks was enough to remove the watermark.

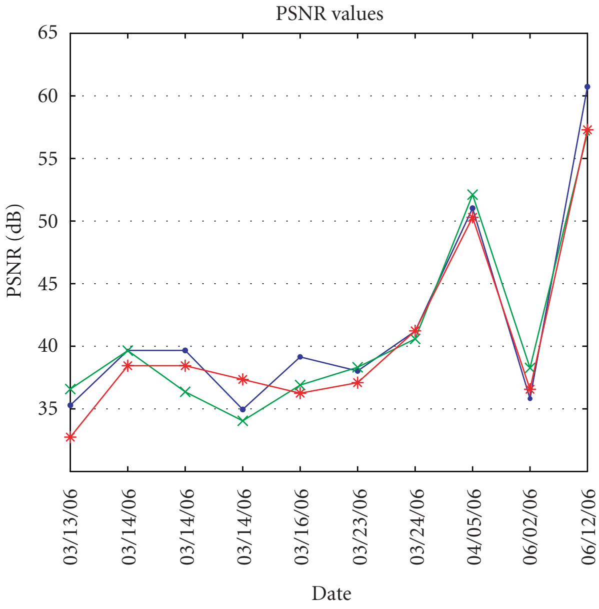


Figure 6: PSNR values between the attacked images and the watermarked ones of the ten best attackers, in chronological ascending order.

were considered: the mean squared error (MSE—basically the same as the PSNR), the mean absolute error (MAE), the maximum absolute error (MAXAE), and the mean structural similarity index (MSSIM) [18]. A precise definition of the above measures is given below; if X and Y represent the images to be compared, we have

$$\begin{aligned} \text{MSE} &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(x(i,j)-y(i,j)\right)^2, \\ \text{MAE} &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|x(i,j)-y(i,j)\right|, \qquad (1) \\ \text{MAXAE} &= \max_{i,j}\left|x(i,j)-y(i,j)\right|. \end{aligned}$$

Table 6: The ten best results of the BOWS hall of fame, reordered according to the different quality measures (averaged over the three images). The attacks carried out in the second phase are the ones with the names in Italic.

Participant (position)	MSE	MAE	MAXAE	MSSIM
A. Westfeld (1)	0.10 (1)	0.064 (1)	3.33 (1)	0.9996 (1)
A. Westfeld (2)	0.51 (2)	0.065 (3)	14.33 (2)	0.9991 (2)
A. Westfeld (3)	5.17 (3)	0.067 (4)	36.67 (7)	0.9984 (4)
Team Craver (4)	7.79 (4)	0.075 (2)	57.67 (9)	0.9982 (3)
Team Craver (5)	10.45 (5)	0.090 (6)	86.67 (10)	0.9978 (6)
A. Westfeld (6)	10.84 (6)	0.098 (8)	125.67 (3)	0.9975 (5)
J. Earl (7)	12.21 (7)	0.099 (5)	206.33 (8)	0.9974 (8)
M. Chekrallah (8)	13.70 (8)	2.019 (10)	216.00 (6)	0.9748 (7)
G. Le Guelvouit (9)	19.46 (9)	2.385 (7)	230.67 (4)	0.9680 (10)
M. Noisternig (10)	22.71 (10)	2.571 (9)	230.67 (5)	0.9514 (9)

The structural similarity index (SSIM) compares local patterns of pixel intensities that have been normalized for luminance and contrast,

$$\text{SSIM}(\mathbf{x},\mathbf{y}) = \left[l(\mathbf{x},\mathbf{y})\right]^{\alpha}\cdot\left[c(\mathbf{x},\mathbf{y})\right]^{\beta}\cdot\left[s(\mathbf{x},\mathbf{y})\right]^{\gamma}, \qquad (2)$$

where $\mathbf{x}$ and $\mathbf{y}$ are subregions of the images X and Y, $l(\mathbf{x},\mathbf{y})$, $c(\mathbf{x},\mathbf{y})$, $s(\mathbf{x},\mathbf{y})$ are respectively the luminance, the contrast and the structure comparison functions, properly weighted by means of the exponents α, β, and γ. The local measures are then weighted obtaining the MSSIM,

$$\text{MSSIM} = \frac{1}{M}\sum_{i=1}^{M}\text{SSIM}(\mathbf{x}_i,\mathbf{y}_i). \qquad (3)$$

Let us note that with MSSIM, greater values indicate greater image quality, while with MSE, MAE, and MAXAE greater values indicate lower qualities.

In Table 6 the hall of fame reordered according to the above criteria is shown. As it can be seen, the ranking changes with the different measures; however, the best result of the contest does not change; if we concentrate on the analysis of the best results of the first phase, the winner would still be S. Craver, with the only exception of the MAXAE, where the best result is achieved by G. Le Guelvouit. These results confirm that no particular differences would have been obtained by using a different quality measure, at least for the winners of the two phases of the contest.

5. CONCLUSIONS

In the framework of the activities carried out by the European Network of Excellence for Cryptology ECRYPT, the BOWS contest was designed to allow to investigate how and when an image watermarking system can be broken though preserving the highest possible quality of the modified content, in case that the watermarking system is subjected to a worldwide attack. During the six months of the contest, about 800 000 images were uploaded into the BOWS server to carry out the attacks on the selected images, coming from more than 300 different IP addresses. The results of the second phase were deeply influenced by the massive use of the sensitivity attack. In any case, we believe that the initiative was a success, and will give many hints to the research in the watermarking area. We then decided to maintain open the BOWS contest website, by adding links to papers related to it, and by allowing interested people to download the attacked images of the best ten contenders, analyzed in the previous section.

In general, the validity of the contest tool to analyze the security and robustness of practical watermarking schemes and to stimulate new research in the area has been widely recognized. We believe that the whole watermarking community will resort more often to this kind of activity in the future (indeed at the moment of writing a second BOWS contest has already been launched, see http://bows2.gipsa-lab.inpg.fr for more information about this initiative).

ACKNOWLEDGMENTS

The work described in this paper has been supported by the European Commission through the IST Programme under Contract IST-2002-507932 ECRYPT. The information in this document reflects only the author's views and is provided as is, and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

REFERENCES

[1] "Ecrypt-european network of excellence for cryptology," http://www.ecrypt.eu.org/, 2004–2008.

[2] A. Kerckhoffs, "La cryptographie militaire," *Journal des Sciences Militaire*, vol. 9, pp. 5–38, 1883.

[3] M. L. Miller, G. J. Doërr, and I. J. Cox, "Applying informed coding and embedding to design a robust high-capacity watermark," *IEEE Transactions on Image Processing*, vol. 13, no. 6, pp. 792–807, 2004.

[4] I. J. Cox, M. L. Miller, and J. A. Bloom, *Digital Watermarking*, Morgan Kaufmann, San Francisco, Calif, USA, 2001.

[5] S. A. Craver, I. Atakli, and J. Yu., "How we broke the BOWS watermark," in *Security, Steganography, and Watermarking of Multimedia Contents IX*, E. J. Delp and P. W. Wong, Eds., vol. 6505 of *Proceedings of SPIE*, p. 65051C, San Jose, Calif, USA, January 2007.

[6] I. J. Cox and J. P. M. G. Linnartz, "Public watermarks and resistance to tampering," in *Proceedings of the 4th IEEE International Conference on Image Processing (ICIP '97)*, vol. 3, pp. 3–6, Santa Barbara, Calif, USA, October 1997.

[7] T. Kalker, J. P. Linnartz, and M. van Dijk, "Watermark estimation through detector analysis," in *Proceedings of the 5th IEEE International Conference on Image Processing (ICIP '98)*, vol. 1, pp. 425–429, Chicago, Ill, USA, October 1998.

[8] P. Comesaña, L. Pérez-Freire, and F. Pérez-González, "Blind newton sensitivity attack," *IEE Proceedings on Information Security*, vol. 153, no. 3, pp. 115–125, 2006.

[9] J. Earl, "Sensitivity analysis for BOWS on the detection region boundary," to appear in *EURASIP Journal on Information Security*.

[10] A. Westfeld, "A workbench for the BOWS contest," to appear in *EURASIP Journal on Information Security*.

[11] P. Comesaña-Alfaro and F. Pérez-González, "Breaking the BOWS watermarking system: key guessing and sensitivity attacks," to appear in *EURASIP Journal on Information Security*.

[12] S. A. Craver, I. Atakli, and J. Y., "Reverse-engineering a watermark detector using an oracle," to appear in *EURASIP Journal on Information Security*.

[13] G. L. Guelvouit, T. Furon, and F. Cayre, "The good, the bad, and the ugly: three different approaches to break their watermarking system," in *Security, Steganography, and Watermarking of Multimedia Contents IX*, E. J. Delp and P. W. Wong, Eds., vol. 6505 of *Proceedings of SPIE*, p. 650517, San Jose, Calif, USA, January 2007.

[14] J. Bennour, J.-L. Dugelay, and F. Matta, "Watermarking attack: BOWS contest," in *Security, Steganography, and Watermarking of Multimedia Contents IX*, E. J. Delp and P. W. Wong, Eds., vol. 6505 of *Proceedings of SPIE*, p. 650518, San Jose, Calif, USA, January 2007.

[15] J. W. Earl, "Tangential sensitivity analysis of watermarks using prior information," in *Security, Steganography, and Watermarking of Multimedia Contents IX*, E. J. Delp and P. W. Wong, Eds., vol. 6505 of *Proceedings of SPIE*, p. 650519, San Jose, Calif, USA, January 2007.

[16] A. Westfeld, "Tackling BOWS with the sensitivity attack," in *Security, Steganography, and Watermarking of Multimedia Contents IX*, E. J. Delp and P. W. Wong, Eds., vol. 6505 of *Proceedings of SPIE*, p. 65051A, San Jose, Calif, USA, January 2007.

[17] P. Comesaña-Alfaro and F. Pérez-González, "Two different approaches for attacking BOWS," in *Security, Steganography, and Watermarking of Multimedia Contents IX*, E. J. Delp and P. W. Wong, Eds., vol. 6505 of *Proceedings of SPIE*, p. 65051B, San Jose, Calif, USA, January 2007.

[18] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.