Attack Model and Performance Evaluation of Text Digital Watermarking

Xinmin Zhou, Sichun Wang and Shuchu Xiong
Information School, Hunan University of Commerce, Changsha, China
Email: xmzhou2008@gmail.com

Jianping Yu
College of Mathematics and Computer Science, Hunan Normal University, Changsha, China
Email: jianpinghn@163.com

***Abstract*—In this paper an attack model of text watermarking is proposed based on communication model, and three different assumptions are made to attackers' actual ability. According to the related research result of digital watermarking theory, the important watermarking properties which influence watermarking robustness and security are analyzed, and then text watermarking robustness and security are evaluated. This work will further propel the development of text watermarking, promote the improvement of text watermarking algorithms and provide some guidance and help to design secure text watermarking schemes.**

***Index Terms*—text digital watermarking, attack model, performance evaluation, information hiding, network security**

I. INTRODUCTION

Earlier study on digital watermarking is only a tentative work, mainly focus on the study of various watermarking algorithms. However, with the rise in watermarking algorithms, the research trend of digital watermarking has begun to change. With the deeper research on digital watermarking, much important research progress is reflected in the increasingly accurate watermarking model since the late 1990s.

The research of digital watermarking model benefits from the thought of communication systems. In essence, digital watermark can be seen as a form of communication. In watermarking scheme, the watermark information is sent from the watermark embedder to the watermark detector in order to achieve communication. Therefore, many experts and academics study the watermarking system according to the conventional communication system model. The study of communication system model which transmitter carries side information was initially launched by Shannon. In this model, the embedder can utilize some information on the channel noise, especially some information on carrier itself. Since Shannon introduced the theory, more and more researchers started the study on the communications with side information [1]. In this theory, for certain types of channels, whether the sender and the receiver are able to receive side information is not important, the most important thing is to eliminate its interference. Recently, some researchers have started to apply this theory of communications with side information to the study of digital watermarking and proposed the watermark models based on communications with side information [2, 3, 4].

In the reference [5], three communication-based models of watermarking are introduced: the basic watermarking model, the watermarking model as communications with side information and the watermarking model as multiplexed communications. In the basic model, the cover Work is considered purely as a noise. In the second model, the cover Work is still considered as a noise, but this noise is provided to the channel encoder as side information. The third model does not consider the cover Work as a noise, but rather as a second message that must be transmitted along with the watermark message in a form of multiplexing. The differences between these models lie in how they incorporate the cover Work into the traditional communications model.

Compared with the research on image watermarking, audio watermarking and video watermarking, there is a great deal of difficulties in the research on text watermarking. However, Inspired by those principles and thoughts of other multimedia watermarking, the study of watermarking technique suitable for text documents can be fueled to some extent. Currently, most research focus on the study of text watermarking algorithms, but less on the research of performance analysis and evaluation of text watermarking. In order to properly evaluate the performances of text watermarking schemes and to draw a fair comparison between different schemes, performances of watermarking schemes should be analyzed and evaluated under comparable conditions. Therefore, it is a very important research to establish reasonable evaluation criteria for text watermarking. In this paper, the communication-based attack model of text watermarking and the theory on performance evaluation of text watermarking are presented, which can provide some convenience for theoretical research and applied research in this field.

doi:10.4304/jcp.5.12.1933-1941

The rest of this paper is organized as follows: Section 2 presents the communication-based attack model of text watermarking, and then the known attacks under the three given conditions are analyzed. The accidental watermarking performances, detection error (reliability), capacity and imperceptibility, will be analyzed in Section 3, and the important performances, watermarking robustness and security will be evaluated in Section 4. Finally, Section 5 concludes the paper.

II. WATERMARKING ATTACK MODEL

Inspired by the thoughts of communication model mentioned above, a better description of attack model on text watermarking is given here, we represent the watermarking attack model as *WAM*, where

WAM=< *{C, M, W, K, C_w, W′}, {A, C_w′, WatAlg, WatDet, WatPar}, {Gen, Emb, Det, Att}* >

In order to facilitate the understanding of this model, some variables are described as follows:

C: watermark carrier;
C_w: watermarked carrier;
C_w′: attacked watermarked carrier;
M: watermarking message;
W: watermark;
W′: extracted watermark;
A: attack;
K: secret key;

K^{gen}, K^{emb} and K^{det} represent the set of secret keys used in the process of watermarking generating, embedding and detection, respectively. In a set, the small letter represents an element of the set of the corresponding capital letter, for example, $C = \{c|c\in\{c_1, c_2, \ldots,\}\}$. Here write algorithms $O\leftarrow$Alg(I) to denote running Alg on inputs I and assigning the output to variable O. Optional inputs or outputs are set in squared brackets, i.e., in Alg(I_1, [I_2]) the input of I_2 is optional, and [I_1, I_2] represents that we have the alternative of choosing the input of I_1 or I_2.

A fundamental attack model on text watermarking is illustrated in Figure 1, the dotted line represents that the corresponding input or output is optional. In order to illustrate the model conveniently, the attack come from the adversary is regarded as the additive noise simply. The watermarking attack model consists of four main parts: watermarking generating *Gen*(), watermarking embedding *Emb*(), watermarking detection *Det*() and watermarking attack *Att*(), which are described as follows, respectively.

$$Gen: W \leftarrow Gen([C], M, [K^{gen}]) \quad (1)$$

$$Emb: C_w \leftarrow Emb(C, W, K^{emb}) \quad (2)$$

$$Det: ([W'], [yes, no]) \leftarrow Det(C_w', K^{det}, [W]) \quad (3)$$

$$Att: A \leftarrow Att([WatAlg], [WatDet], WatPar) \quad (4)$$

The processes of watermarking generating, embedding and detection are similar to general watermarking model, so these processes are not repeated here.

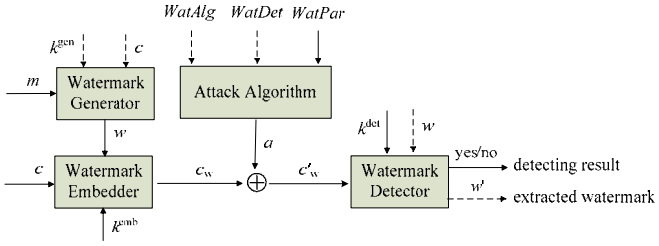

Figure1. Attack model of text watermarking

However, watermarking attack is the most important part in watermarking attack model, in which three attack abilities of adversaries are considered. The process of watermarking attack can be simply denoted as follows:

$$A \leftarrow Att([WatAlg], [WatDet], WatPar) \quad (5)$$

$$C_w' \leftarrow C_w + A \quad (6)$$

where C_w' is a set of watermarked text documents may be attacked by adversaries. *WatAlg* denotes watermarking algorithm; *WatDet* denotes watermarking detector; *WatPar* denotes watermarking parameter, such as *C*, *M*, *W*, *K* and C_w, it can be any combination of them and can be also empty. However，the adversary can obtain other processing software and algorithms which may be unknown, these tools will increase the difficulty of analysis on watermarking attack.

In this Section, we will discuss the attack analysis based on the watermarking attack model mentioned above. Generally, it is extremely rare for the attacker knows nothing or all about the watermarking scheme. Depending on the Kerckhoffs' assumption of the cryptographic community, the following discussions are based on the assuming that the adversary knows nothing about the secret keys. Here we analyze the known attack in the context of the three given conditions, respectively. Three given conditions are defined as follows: $cond_1$={C_w}, $cond_2$={*WatAlg*, *WatPar*}, $cond_3$={*WatDet*, *WatPar* }. The following expressions $C_w' \odot C_w$ and $C_w' \oplus C_w$ denote C_w' does contain and does not contain the watermark, respectively.

1. $cond_1$={C_w}: If an adversary obtained multiple watermarked Works, the adversary can often exploit these Works to remove watermarks, even if he or she knows nothing about the algorithm. Usually, attacks that rely on possession of several watermarked Works are known as collusion attacks. This successful attack process can be represented by the following expressions:

$$A \leftarrow Att(C_w) \quad (7)$$

$$C_w' \leftarrow C_w + A \quad (C_w' \oplus C_w) \quad (8)$$

$$no \leftarrow Det(C_w', K^{det}, [W]) \quad (9)$$

2. $cond_2$={*WatAlg*, *WatPar*}: An adversary who has complete knowledge of watermarking algorithms can find and exploit weaknesses in algorithms. Any process that maintains the fidelity of the watermarked Work could be used by an adversary to identify specific distortions for which the detector cannot compensate, and then apply a successful masking attack, or eliminate the watermark. This attack process can be represented by the following expressions:

$$A \leftarrow Att\ (WatAlg,\ WatPar) \quad (10)$$

$$C_w' \leftarrow C_w + A \quad (C_w' \odot C_w \text{ or } C_w' \oplus C_w) \quad (11)$$

$$\text{no} \leftarrow Det\ (C_w', K^{det}, [W]) \quad (12)$$

For a secure watermark, it must be robust to any process that maintains the fidelity of the Work. Otherwise，once the adversary has gained the secret of watermarking algorithms, he might be able to perform unauthorized embedding which is represented by the following equation, where K^{emb*} denotes the illegal embedding key.

$$C_w^* \leftarrow Emb(C, W, K^{emb*}) \quad (13)$$

3. $cond_3$ ={*WatDet*, *WatPar* }: Even if an adversary does not know anything about the watermarking algorithms, access to a watermark detector will give him a great advantage in attacking the watermark. By making iterative modifications to the watermarked Work, and testing after each change, the modified watermarked Works tend to fall into two primary categories: C_w^1 and C_w^2, the detecting processes of them are presented as following expressions:

$$([W'], yes) \leftarrow Det\ (C_w^1, K^{det}, [W]) \quad (14)$$

$$no \leftarrow Det\ (C_w^2, K^{det}, [W]) \quad (15)$$

By observing the detector's results, the adversary can learn a great deal about how the detector operates, and the obtained knowledge can be exploited in the sensitivity analysis attack. This attack process can be represented by the following expressions:

$$A \leftarrow Att\ (WatDet,\ WatPar) \quad (16)$$

$$C_w' \leftarrow C_w + A \quad (C_w' \oplus C_w) \quad (17)$$

$$no \leftarrow Det\ (C_w', K^{det}, [W]) \quad (18)$$

III. Performance Analysis of Text Watermarking

Watermarking systems can be characterized by a number of defining properties. The relative importance of each property is dependent on the requirements of the application and the role the watermark will play. In fact, even the interpretation of a watermark property can vary with the application. These watermarking properties can be used as performance criteria to evaluate watermarking schemes and provide some favorable guidance for the design of watermarking schemes with certain application background. In this Section, some watermarking performances, such as detection error, capacity and imperceptibility which affect watermarking robustness and security to some extent are analyzed here in detail. Bit error rate (BER) can also be used as an important parameter to distinguish the good from the bad of watermark systems. However, considering the potential interference from the communication channel, BER is not a true indication of watermarking robustness when it is measured by BER Merely. In the research on watermarking algorithms, watermarking robustness and security can be used to distinguish watermarking algorithms between good and bad. We will evaluate these two important performances in next Section.

A. Detection Error Analysis

The design of a good watermarking system should consider three aspects: watermark generating strategy, embedding strategy and detecting strategy. Generalized watermarking detection refers to the watermark detector should be able to determine whether the cover contains a watermark, and be able to extract the complete watermark correctly by an appropriate method. Detection error refers to the watermarking detector made an error during determining the existence of watermark. Correlation detection is a common strategy of watermark detection. In the correlation detection, the detecting threshold setting played a decisive role on the detecting results. This detection can be regarded as a binary hypothesis validation, and the two main types of detection error are presented as follows:

Error I: False Positive Error, which occurs when the detector incorrectly indicates that a watermark is present;

Error II: False Negative Error, which occurs when a detector incorrectly indicates the absence of a watermark.

It's worth noting that errors are inevitable in even the best-designed watermarking systems. In addition to these two types of errors mentioned above, a message error occurs when a watermark detector incorrectly decodes a message. The false negative probability is highly affected by the distortions the Work undergoes between the times of embedding and detection. Whereas false positives depend only on the detection algorithm, false negatives also depend on the embedding algorithm. False positive and false negative errors are coupled. Figure 2 [5] illustrates how and why false positive errors can occur. It is more intuitive to see the resulting situation of these two categories of errors. In Figure 2, the left-hand curve represents the frequency of occurrence of each possible value that can be output from the watermark detector when no watermark is actually present. Similarly, the curve to the right represents the frequency of detector output values when a watermark is present. The vertical dotted line represents the decision boundary τ. The specific decision rule is presented as follows: If the detector output value is less than τ, the watermark is declared absent; otherwise, the watermark is declared present. Usually, the output of the detector is the similarity between the extracted watermark and the original watermark. Currently familiar formulas called normalized correlation, which are used for evaluating the similarity degree of digital watermark, showed a big shortage when expressing the similarity degree of extracted watermark and original watermark in inverse similarity. Therefore, the reference [7] proposed a polarized correlation method to evaluate the similarity degree of digital watermarking. Experimental results show that the method is more scientific and reasonable. In Figure 2, the shaded area A underneath the curve represents the probability of a false negative, and the shaded area B represents the probability of a false positive. Usually, the bigger the detecting threshold, the bigger the false negative probability; otherwise, the false positive probability is bigger.

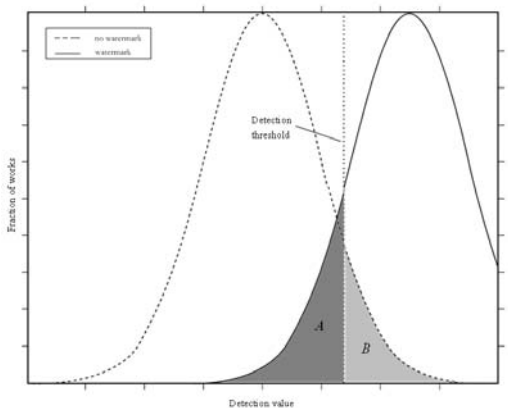

Figure 2 Detector output distribution and detection threshold

B. Watermarking Capacity Analysis

Digital watermarking system is essentially a communications system, which transmits information from the watermark embedder to the watermark receiver, the watermark itself is the transmitted information of the system and the cover object is regarded as the channel. It is natural, then, to try to fit watermarking into the traditional model of a communications system, and use communication theory to analyze performances of watermarking system. Watermarking capacity analysis is a key issue of digital watermarking technique. In the study on watermarking algorithm, it is very important to know how much information can be reliably embedded in a given carrier signal. Currently, there is more literature on capacity analysis of digital watermarking. In the reference [8], the diversified research results of channel capacity of digital watermarking are summarized, and the capacity of watermarking as a basic communication system, as side information, using dirty paper encoding and security watermarking are introduced. By analyzing the capacity, an effective conclusion for the designing of embedding algorithm and detecting algorithm is found in this paper. Information theory pointed out that any discrete channel interference, if X represents the transmitted information of channel, Y represents the received signal, and then this channel capacity can be expressed as follows:

$$\begin{aligned} C &= \max_{P_X(x)} I(X,Y) \\ &= \max_{P_X(x)} [H(Y) - H(Y \mid X)] \\ &= \max_{P_X(x)} [H(X) - H(X \mid Y)] \end{aligned} \tag{19}$$

In which, P_X (x) denotes the probability distribution function of the signal X; I(X, Y) denotes the mutual information between X and Y; H (Y) denotes the entropy of the signal Y; H(Y |X) denotes the conditional entropy, namely, the expectation entropy of Y when X was determined. The basic digital watermarking model is similar to the general discrete noise channel, the channel capacity C of basic spread spectrum watermarking system can be calculated based on the Equation 19 mentioned above. The specific expressions [9] are presented as follows:

$$C_1 = \frac{1}{2}\log_2[1 + \frac{\sigma_W^2}{\sigma_N^2}] \tag{20}$$

$$C_2 = \frac{1}{2}\log_2[1 + \frac{\sigma_W^2}{\sigma_X^2 + \sigma_N^2}] \tag{21}$$

where, assuming the original image X follows the Gaussian distribution (0, σ_X^2) with the mean 0 and the variance σ_X^2 , the attack noise N follows the Gaussian distribution (0, σ_N^2) with the mean 0 and the variance σ_N^2 and the watermark signal W follows the Gaussian distribution (0, σ_W^2) with the mean 0 and the variance σ_W^2 . The mentioned-above Equation denotes the channel capacity of spread spectrum watermarking with additive Gaussian white noise. C_1 and C_2 denote watermarking capacity of non-blind watermarking algorithm and blind watermarking algorithm, respectively. for blind and watermark capacity watermarking algorithm. As the non-blind watermarking algorithm is able to use information of original cover during extracting watermark, the watermark capacity of blind watermarking algorithm is often less than the watermark capacity of non-blind watermarking algorithm.

The original carrier information is neglected in the basic watermark model during the watermark encoding and decoding, but the watermark model with side information [10, 11] during the process of watermark encoding and decoding regards the original carrier signal as side information to reduce the impaction of the original carrier on the watermarking system performance. Such type of channel capacity is explained in reference [12] and is denoted as follows:

$$C = \max_{P_{SUX}} [I(U;Y) - I(U;S)] \tag{22}$$

Where S is the known value of side information of the transmitter, which is independent of the encoding method; U is an auxiliary variable, it is determined by the distribution of message m and the impact of S on the encoding method: X is an element of the transmitted signal, generally, it is a function of S and U, which is X = f (U, S), P_{SUX} is the joint probability distribution of the three variables of S, U and X.

It can be seen from the analysis of channel capacity of watermarking system that channel capacity of watermark system is closely related to watermarking noise. The overall trend is that the signal noise ratio is increased, the watermarking channel capacity increases monotonically. Using the side information of the communication model to analyze the channel capacity, on the whole, the channel capacity of this watermarking system is better than the watermarking capacity based on the communication model.

C. Analysis of invisibility

Watermark embedding will result in the decrease of visual quality of watermark cover. Therefore, during the design and analysis of watermarking algorithm, watermark invisibility should be considered fully. Two quantitative measurement methods for visual measurement of watermark cover quality are presented: one is the pixel-based metrics; the other is the visibility

quality metrics. In general, the watermark embedding will lead to the change in different degree of distortion measurement. Therefore, it is not accurate that evaluating the change degree of watermark cover only by the pixel-based measurement method. Usually, it is more effective to use the visibility quality metrics which appropriate for the human visual system to evaluate the change degree of watermark cover.

Visibility quality measurement exploited the contrast sensitivity and shielding phenomenon of the human visual system. The calculation steps of measurement are presented as follows: Firstly, block the image and decompose the coding error and the original image to the various perceptual components with the filters; Secondly, calculate the detection threshold of each pixel of the image, and then calculate the filtering error according to the given threshold; Finally, all the color channels are made the upper operations, the difference above the threshold value is the consistency measurement, namely, Just Noticeable Difference (JND). The measuring formula is MPSNR (Masked Peak Signal to Noise Ratio), and the computation formula is presented as follows:

$$MPSNR = 10\log_{10}\frac{255^2}{E^2} \tag{23}$$

Where, E denotes the calculation distortion. Because the meaning of this measurement and the known dB is different, so it is represented by visual dB (Visual Decibels, vd_b). The overall quality measurement of the image can be expressed as the quality scale $Q = 5 / (1 + N * 5)$, in which N is a standardized constant, which is usually selected as a value that can make the reference value to map the corresponding value of the quality partition. The specific criteria to measure visibility quality are presented as following Table 1 [13].

Table 1 Quality ratings on a scale from 1 to 5

Rating	Impairment	Quality
5	Imperceptible	Excellent
4	Perceptible, not annoying	Good
3	Slightly annoying	Fair
2	Annoying	Poor
1	Very annoying	Bad

IV. Performance Evaluation of Text Watermarking

At present, as the lack of uniform of watermarking benchmarks, watermarking performances of different watermarking schemes can not be analyzed and compared. In order to complete a fair performance evaluation, testing performance criteria, performance standards, attacking methods and testing objects all need to be unified, and a standard testing procedure is necessary too. During the watermark testing, Petitcolas FAP et al. regarded the whole process of watermark as a black box, only the standard testing parameters are inputed, we can observe the output of performance criteria under various inputs. The specific testing process in this literature is presented as follows [13]:

1) embed watermark with the greatest intensity, ensuring the premise of visual quality;

2) make a series of attacks on watermarked images;

3) extract the watermark and judge the succeed or not for different attacks;

4) repeat the watermark embedding and the mentiond-above process for each watermarked images.

Currently, there are three main types of representative watermark testing benchmarks: 1) Stirmark: proposed by Petitcolas FAP during his studies towards the doctoral degree at Cambridge University, in november 1977 the first version was published, now it has become the most widely used evaluating tools of watermark robustness; 2) Checkmark: proposed by the study group of computer vision leaded by professor Thierry Pun of the University of Geneva, the initial version was published in June 2001, Checkmark considered attacks which no mentioned in Stirmark, and during the evaluation of attacks by this benchmark tool, the specific watermarking applications were taken into account; 3) Optimark: proposed by Artificial Intelligence and Information Analysis Laboratory of information department in Aristotle University, Greece. It is a testing benchmark tool for static image watermarking algorithms. The difference between Stirmark and Checkmark is, Optimark has a graphical interface, different watermark keys and information can be used for performance evaluation by multiple testing of detecting and decoding. The functions and attack types of three types of watermark benchmark tools mentioned above were analyzed and compared in detail in the reference [14].

If the good and consistent benchmark can not be used to evaluate performances of watermarking systems, the watermarking system's vulnerability would been existed all the time, and it is more difficult to compare the pros and cons of various watermarking systems. This is a serious impediment to the research and application of watermarking technique. Based on these considerations, in March 2000, Certimark (Certification for Watermarking Techniques) program of European Union was started, which goal is to establish international uniform standards to ensure the standardization of watermarking testing [14]. Certimark was designed to provide possible complete benchmarks for still images, video technology and the future multimedia system; to develop the related tools for these benchmarks in order to protect the security of multimedia transmitted in Internet; to mingle and integrate all testing parameters and modular integration of different digital watermarking algorithms, attacks and video quality assessment are allowed; to study high-end digital watermarking algorithm, and the most promising watermarking technique is applied to evaluation benchmarks of Certimark.

The emergence and research of watermark benchmark tools mentioned above provide a great advantage for watermarking performance evaluation and analysis. In this section, borrowing the thought from the related

research findings of image watermarking, robustness and security evaluation on text watermarking are analyzed and summarized to provide some help and guidance for the research on text watermarking in future.

A. Descriptions of Watermark Performance Evaluation

For a reasonable performance evaluation, a good testing environment should be controlled, that is to say, some parameters should be fixed. Some useful charts, variables and constants can be used to be compared are given in Table 2 [13].

Table 2 Different graphs and corresponding variables and constants

Graph type	Parameter			
	Visual quality	Robustness	Attack	Bits
Robustness vs attack	fixed	variable	variable	fixed
Robustness vs visual quality	variable	variable	fixed	fixed
Attack vs visual quality	variable	fixed	variable	fixed
ROC	fixed	fixed	fixed/ variable	fixed

Experimental conditions: the testing object is the color Lena image of 512 × 512 × 24, the embedded watermark length is 100 bits and JPEG compression attack is used, the key is the seed of random number generator used to generate the spread spectrum sequence; bit error rate is used as the index of watermark robustness evaluation, the visual quality metrics is represented by the quality score Q of the visibility quality metrics; two main watermarking methods were compared: the method of spatial domain watermark embedding and the method of watermark embedding in multi-resolution environment.

According to the testing environment and the experimental conditions in Table 2, the following curves [13] could be obtained by a large number of experiments. The specific roles of these curves in watermark performance evaluation are described as followed:

1) curve of robustness vs attack: this curve reflects the functional relationship between bit error rate and attack power given the visual quality. watermark Robustness can be directly compared by this curve, which can show the overall robust performance for attack. The curve indicates that given visual quality, performance of multi-resolution watermarking scheme is better. (Q = 4.5, each test using a different key and the testing times 10)

2) curve of robustness vs visual quality: this curve reflects the functional relationship between bit error rate and visual quality given attack power.The curve can be used to obtain a minimum of visual quality on the specific request of attack power and bit error rate. This curve shows that the multi-resolution watermarking scheme has a higher visual quality for given expected bit error rate. (Using the JPEG compression attack of the quality factor 75%)

3) curve of attack vs visual quality: this curve reflects the functional relationship between maximum allowable attack power and visual quality for given robustness. The curve can give a direct evaluation of allowable attack power under the premise of a given visual quality. The curve can compare robustness between different watermarking methods on the requirement for given bit error rate and visual quality. (Each test using a different key, 5 times, bit error rate 0.1)

4) ROC curve: this curve play a very important role in evaluating the performance and reliability of the watermarking scheme during watermark detection. Usually, the detector performance is better, its ROC curve is more close to the upper left corner, the curve integration can be used as the measured index of watermark detector performance. (Each test using a different key, 10 times, Q= 4.5, using JPEG compression attack, the quality factor changes from 30% to 100% with the interval of 5%)

ROC curve reflects the functional relationship between TPF (TPF: True Positive Fraction) of y-axis and FPF (FPF: False Positive Fraction) of x-axis. TPF is defined as follows:

$$TPF = \frac{TP}{TP + FN} \tag{24}$$

In which, TP and FN denote true positive times and false negative times of testing results, respectively. FPF is defined as follows:

$$FPF = \frac{FP}{TN + FP} \tag{25}$$

Where FP and TN denote false positive times and true negative times of testing results, respectively. It is important to note that FPF can be regarded as the corresponding probability of false positive, but the corresponding probability of false negative is equal to 1 minus TPF.

B. Watermark Robustness Evaluation

In the research on digital watermarking, robustness and imperceptible are two most important indexes of evaluating watermark embedding algorithms, and the research on robust watermarking algorithm is the chief content of the research on digital watermark [16, 17]. Two key factors affecting watermark robustness are: watermark structure and embedding strategy [18]. The current research on digital watermark mainly focuses on how to embed and extract watermark, while less on the watermark itself structure and its properties.

Embedded watermarks could possibly have many different forms, which can be text or ID (Identification), graphic, image, audio and other random sequence. Currently, the most research focus on random sequence watermark, but the research on meaningful watermark has attracted the attention. There is no doubt that meaningful watermark is more intuitive and verifiable than random sequence watermark in copyright protection and content authentication. However, in order to make watermark more robust, most watermarking algorithms use pseudo-random sequence (Gaussian Sequence, Uniform Sequence, and Binary Sequence) as watermarks. Meanwhile, the use of error correction encoding can further improve the watermark robustness and indirectly

play a role of encryption. Thus, the analysis of watermark structure and property will be favorable for determining the best form of watermark. This is of great significance to the research on watermark embedding algorithms.

It is an important part of the study on watermark establishing a reasonable evaluating method and a testing benchmark for watermarking performance evaluation. The evaluation of watermark robustness mainly considers two aspects: watermark robustness and transparency. In general, the design of watermarking system should compromise between robustness and transparency. Therefore, in order to carry out a fair and reasonable performance evaluation, it should be ensured that all the watermarking systems must be tested under comparable conditions. The analysis and discussion of main factors impacting watermark robustness based on this testing premise are presented as followed [13]:

1) embedding capacity: because it directly affects the watermark robustness, it is an important parameter of robustness. For the same kind of watermarking scheme, the more the watermark information embedded, the lower the watermark robustness. The embedded information is dependent on different applications.

2) embedding intensity: there is a compromise between watermark embedding intensity and watermark visibility. The enhancing of watermark robustness will increase the watermark embedding intensity, which will correspondingly increase the watermark visibility.

3) the size and type of data: the size of data carrier has a direct impact on the watermark robustness. In addition to the size of data carrier, the data type also has an important impact on the watermark robustness.

4) secret key: Although the number of secret key does not directly affect the watermark visibility and robustness, it plays an important role in the security of watermark systems. In the design of watermarking algorithms, the key space must be large enough to make exhaustive attack lapsed. Many watermark systems can not resist some simple attacks, because their design does not follow the basic principle of cryptography.

Each performance of watermarking systems mutually supports and constrains with each other [19]. By the secret key, the embedding location and intensity of watermark are difficult to be assumed, which make the attacker can not easily destroy the watermarked Work for the price of the decrease of perceived quality, thereby it supports the implementation of the robustness; and the realization of blind detection not only makes watermarking applications more convenient, but also enhances the security of original data. However, the perceived watermark transparency generally required the watermark signal intensity to be small, which affects the residual capacity of the watermark, thus constrains the watermark robustness; high information loads requires reducing the carrying amount of data needed by per unit of watermark information, reducing the redundancy of watermark information, which make the watermark is more sensitive to errors. Obviously, it is the core of the research and application of robust watermark to achieve and balance watermarking performances mentioned-above.

C. Watermark Security Evaluation

At present text watermarking schemes is on the increase. How to improve the security of text watermarking schemes and meet user particular requirements from a practical standpoint is a problem that should be exigent to be solved. Generally, the watermarking security should be considered from the following aspects: first, evaluating watermarking security, whether or not security vulnerabilities exist in watermarking schemes should be judged, and corresponding reliable technical solutions should be found; second, developing security policy and implementing it in watermarking system and network architecture; third, designing security architecture, during the design of security architecture, user particular requirements should be given full considerations. Security evaluation of text watermarking is a very complicated task, which involving manifold factors. In order to improve watermarking security effectively, we should establish relevant security policy based on the different purposes of text watermarking systems by exploiting the principle of stressing focal points.

1) performance evaluation of preventing detection: Determine whether the watermarked carrier can be perceived by human senses, and whether the distinction between the watermarked carrier and the original carrier can be found accurately; Determine whether the leak of watermarking algorithm can be found by the method of statistical analysis, and whether it has the ability of preventing statistical detection; the changed degree of the watermark carrier is larger, the watermark can be detected more easily, and vice versa. Usually, because text watermarking algorithm based on semantics makes the minimal damage to the context and syntax of the text carrier, it is regarded as the more secure watermarking algorithm.

2) performance evaluation of preventing capacity estimation of watermark: how to determine the maximum of text watermark capacity is an important aspect of security evaluation. During the analysis of text watermarking capacity, firstly, the analysis model of watermark capacity is established based on its watermark embedding algorithm, and then in the corresponding model, increase the watermark capacity until it reaches the theoretical maximum of the secure capacity estimated value. If the maximum of watermark capacity of the evaluated text watermark system is equal to or less than the maximum of secure capacity estimated value of this model, the text watermark system is thought to be secure in preventing capacity estimation of watermark.

3) performance evaluation of preventing watermark removal: to ensure the watermarked text can be extracted legally and achieve the original purpose of the watermark when it is transmitted in the Internet and suffer varying degrees of destruction, text watermark embedding algorithm should be robust to common signal processing and has the capability to resist malicious attacks.

4) performance evaluation of preventing watermark utilization: preventing utilization of text watermarking system is the key distribution mechanism for preventing

the middle attack. In the current text watermarking systems, there are two main key delivery systems: watermarking system of symmetric key (private watermarking system) and watermarking system of asymmetric key (public watermarking system). The security level of public watermarking system is higher than private watermarking system. Because the watermarking system of zero-knowledge proof does not reveal any information about the key, therefore, its security is better. Zero-knowledge proof should attract the adequate attention to develop security strategy in the designing of text watermark system.

Table 3 Security policy of text watermarking system

Application of text watermarking system	The key of developing security policy
Covert communication	Preventing watermark detection
General message communication	Preventing capacity estimate of watermark
Copyright protection	Preventing watermark removal
Electronic Commerce	Preventing watermark utilization

The four security links mentioned above are actually conflicting. For example, when the text authentication requires more high security, the watermark embedding capacity is necessarily required as small as possible; but if the amount of embedded information for text authentication reduced to the amount can not express the transmitted message accurately, it will violate the fundamental purpose of text authentication. Based on the above four indexes for the evaluation of text watermark security, Table 3 shows the key of developing security strategy for different applications of text watermark systems. In order to overcome these contradictions efficiently, we should develop the relevant security strategy by the principle of focusing on key projects, according to the different application of text watermarking systems.

V. Conclusions

In essence, digital watermarking can be regarded as a form of communications. In this paper, firstly we propose the attack model of text watermarking based on communications, in which three assumptions about the adversary are presented: 1) if the attacker has multiple watermarked Works; 2) if the attacker knows the watermarking algorithms; 3) if the attacker has a watermarking detector. Then, based on this proposed attack model, the formal analysis on the known attacks under the three given assumptions is expounded. The work of this paper will fuel the algorithmic improvements and provide a help for ensuring watermarking scheme secure.

The research on performance analysis and evaluation of text watermarking is developed. At first, according to the proposed watermarking mathematical model, the formularized analyses on watermarking imperceptibility, capacity and reliability are presented. These watermarking performances make an impact on watermarking robustness and security to some extent. Therefore, the trade-offs between them should be considered. Then, the performance evaluation of robustness and security are presented in detail. These two performances can be used to distinguish text watermarking algorithms between good and bad. The work of this paper will give a help for further research on text watermarking in future, especially in aspects of performance analysis and evaluation of text watermarking.

Acknowledgment

The authors wish to thank for anonymous referees' useful suggestions. This work was supported in part by the National Natural Science Foundation of China under grant No. 60903168 and the Key Technology Research and Development Program of Hunan Province under grant No. 2009GK3131.

References

[1] C. E. Shannon. Channels with side information at the transmitter. IBM Journal of Research and Development, 1958: 289–293

[2] Y. Steinberg, N. Merhav. Identification in the Presence of Side Information With Application To Watermarking. IEEE Transactions on Information Theory, 2001, 47: 1410-1422

[3] R.J. Barron, B. Chen, G.W. Wornell. The Duality Between Information Embedding and Source Coding With Side Information and Some Applications. IEEE Transactions on Information Theory, 2003, 49: 1159- 1180

[4] C.S. Lu. Towards robust image watermarking: combining content-dependent key, moment normalization, and side-informed embedding. Signal Processing: Image Communication, 2005, 20: 129-150

[5] I.J. Cox, M.L. Miller, J.A. Bloom, J. Fridrich, T. Kalker. Digital Watermarking and steganography, Morgan Kaufmann publishers, Seattle, Washington, USA, 2008.

[6] S.Voloshynovskiy, S. Pereira, V. Iquise, T. Pun. Attack modelling: towards a second generation benchmark. Signal Processing, 2001, 81: 1177–1214

[7] Li Xudong. Analysis and Improvement of Formulas for Evaluating Similarity Degree of Digital Watermarks. Acta Automatica Sinica. 2008, 34(2): 208-210 (in Chinese)

[8] Wang Ying, Li Xianglin. An Overview of the Capacity of Digital Watermarking. Journal of Electronics & Information Technology. 2006, 28(5): 955-960 (in Chinese)

[9] JJ Eggers, R Bauml, B Girod. Digital Watermarking Facing Attacks by Amplitude Scaling and Additive White Noise. 4th Int. ITG Conf. on Source and Channel Coding, 2002: 28-30

[10] B Chen, G W Wornell. Achievable Performance of Digital Watermarking Systems. In: Proceedings of the 1999 6th International Conference on Multimedia Computing and Systems. CA: IEEE, Los Alamitos, 1999, 1: 13-18

[11] I J Cox, M L Miller, A Mckellips. Watermarking as Communications with Side Information. Proceeding of the IEEE, 1999, 87(7): 1127-1141

[12] SI Gel'Fand, MS Pinsker. Coding for Channel with Random Parameters. Problems of control and information theory, 1980, 9(1): 19-31

[13] Kutter M, Petitcolas FAP. Fair Evaluation Methods for Image Watermarking Systems. Journal of Electronic Imaging. 2000, 9(4): 445-455
[14] Wang DaoShun, Liang JingHong, Dai YiQi, Luo Song, Qi DongXu. Evaluation of the Val idity of Image Watermarking. Chinese Journal of Computers. 2003, 26 (7): 779-788 (in Chinese)
[15] Petitcolas FAP. Anderson RJ. Evaluation of copyright making systems. Proceedings of IEEE Multimedia Systems'99. 1999: 574-579
[16] Huang Jiwu, Tan Tieniu. A Review of Invisible Image Watermarking. Acta Automatica Sinica. 2000, 26 (5): 645-655 (in Chinese)
[17] Feng Tao, Wang ChengFa, Han JiQing. Research on the Property of Robust Watermarking Structure. Chinese Journal of Computers. 2004, 27(7): 971-976 (in Chinese)
[18] I.J. Cox, J. Killian, F.T.Leighton et al. Secure spread spectrum watermarking for multimedia. IEEE Transactions on Image Processing. 1997, 6(12):1673-1687
[19] Petitcolas FAP. Watermarking Schemes Evaluation. IEEE Signal Processing Magazine, 2000, 17(5): 58-64

XinMin Zhou received his M.S. degree in computer application technology from Hunan University, Hunan, China, in 2006, and received his Ph.D. degree in computer application technology from Tongji University, Shanghai, China, in 2010. He has been a lecturer of Information Department, Hunan University of Commerce, Hunan, China, since 2008. His research interests include text watermarking, information hiding and network security.

SiChun Wang received his M.S. degree in computer application technology from the Chinese Academy of Sciences, Beijing, China, in 1998, and received his Ph.D. degree in control theory and engineering from Central South University, Hunan, China, in 2005. He is the president of Information School, Hunan University of Commerce, Hunan, China. His research interests include decision theory and optimization, network security.

ShuChu Xiong received his B.A. degree in computer application technology from Central South University, Hunan, China, in 1996. He is the dean of the department of Engineering Management, Hunan University of Commerce, Hunan, China. His research interests include information resources management, service science and network security.

JianPing Yu received his B.A. degree in computer application technology from Hunan University, Hunan, China, in 2003, and received his Ph.D. degree in computer application technology from Hunan University, Hunan, China, in 2008. He has been a lecturer of College of Mathematics and Computer Science, Hunan Normal University, Hunan, China, since 2009. His research interests include wireless sensor networks and network security.