Signal Processing: Image Communication 46 (2016) 60–75

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

Identifying deficits of visual security metrics for images

Heinz Hofbauer*, Andreas Uhl

University of Salzburg, Department of Computer Sciences, Austria

ARTICLE INFO

Article history:
Received 26 August 2015
Received in revised form
10 March 2016
Accepted 3 May 2016
Available online 4 May 2016

Keywords:
Confidence
Image metrics
Partial encryption
Security metrics
Selective encryption
Sufficient encryption
Transparent encryption

ABSTRACT

Visual security metrics are deterministic measures with the (claimed) ability to assess whether an encryption method for visual data does achieve its defined goal. These metrics are usually developed together with a particular encryption method in order to provide an evaluation of said method based on its visual output. However, visual security metrics themselves are rarely evaluated and the claim to perform as a visual security metric is not tied to the specific encryption method for which they were developed. In this paper, we introduce a methodology for assessing the performance of security metrics based on common media encryption scenarios. We systematically evaluate visual security metrics proposed in the literature, along with conventional image metrics which are frequently used for the same task. We show that they are generally not suitable to perform their claimed task.

1. Introduction

The claim of visual security metrics (security metrics for brevity) is usually the ability to assess the functionality of an encryption method based on the output of the encryption of visual data. In particular, the evaluation of an encryption method is only based on the visual output (i.e., the ciphertext), which is either an image or video. While such metrics are often created in conjunction with a specific encryption method and tested, if at all, only for this encryption method, the claim to perform as a security metric is usually universal. Furthermore, regular image quality metrics, such as the frequently used PSNR and SSIM, are also utilized in the literature to evaluate encryption methods [1–4].

The problem with the evaluation of security metrics is the fact that there is no established testing methodology. Thus, even if security metrics are tested, the test is usually based on the evaluation procedures for regular image metrics, which are not sufficient to establish whether a method is applicable in the context of encryption.

Regarding cryptographic security, Shannon's work [5] shows that the highest level of security is reached when applying a secure cipher to a redundancy free plain text. Current image/video codecs exploit redundancy for compression and thus we can consider a bit stream to be an almost redundancy free plain text in the sense of Shannon. Consequently, for maximal security, the encryption of the entire bit stream with a state-of-the-art cipher, such as AES, would suffice ("conventional encryption"). Lookabaugh and Sicker [6] showed that selective encryption is sound and demonstrated its relation to Shannon's work. However, [7] showed that side information can compromise security.

However, there are application scenarios which make it necessary to move away from full encryption. Methods which do not utilize full encryption of the underlying data are called Lightweight/Soft/Partial/Selective Encryption. Specifically, selective encryption is the application of an assumed secure encryption method to a *selected* part of the plain text. In selective encryption the *encryption* part is assumed to be secure, e.g., by using AES. The final security of the selective encryption comes then from the *selection* part. What is evaluated in order to gauge the final security of the selective encryption is, to what extent the information left in plain text can be used to reconstruct an image or video.

Furthermore, an attack on the selective encryption method does not come from attacking the *encryption*, but from attacking the *selection*. This is usually done by using knowledge about the original format of the image/video. An attack is usually based on removing the negative impact on quality by the, essentially random, signal introduced by the encryption. This is typically done by replacing the random signal by a signal which introduces the least amount of error into the final decoding. In order to do so, very specific knowledge about the containing format has to be exploited, and there is usually only a single method to go about this, i.e., *the attack*.

* Corresponding author.
E-mail addresses: hhofbaue@cosy.sbg.ac.at (H. Hofbauer), uhl@cosy.sbg.ac.at (A. Uhl).

http://dx.doi.org/10.1016/j.image.2016.05.001
0923-5965/© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Another, (implicit) assumption about the selective encryption method under test is the *format compliance*. Format compliance requires an encrypted bitstream to be decodable by a standards compliant decoder. In other words, format errors may not be introduced by the encryption. Thus, in the following, when we refer to an encrypted image/video we mean the image or video that results from decoding an unattacked encrypted bitstream.

The notion of security in selective encryption is different from the traditional notion of security: First, we knowingly leave information in plain text to retain format compliance; second, the focus is on content security not information security, i.e., the content should be secure (to some defined extent), while information about the content might be allowed to leak. In order to be able to discuss the exact notion of security in such non-conventional encryption schemes, we need to distinguish distinct application scenarios of encryption schemes for visual data:

Confidentiality encryption: Means MP security (message privacy). The formal notion is that if a system is MP-secure an attacker cannot efficiently compute any property of the plain text from the cipher text [8]. This can only be achieved by the conventional encryption approach.

Content confidentiality: This is a relaxation of confidential encryption. Side channel information may be reconstructed or left in plaintext, e.g., header information, packet length, but the visual content must be secure in the sense that it must not be intelligible/discernible [9].

Sufficient encryption: Means we do not require full security, just enough security to prevent abuse of the data. The content must not be consumable due to high distortion (e.g., for DRM systems) by destroying visual quality to a degree which prevents a pleasant viewing experience or destroys the commercial value. This implicitly refers to message quality security (MQ), which requires that an adversary cannot reconstruct a higher quality version of the encrypted material than specified for the application scenario [10].

Perceptual/transparent encryption: Means we want consumers to be able to view a preview version of the video but in a lower quality while preventing them from seeing a full version. As an example: this can be used in a pay per view scheme where a lower quality preview version is available from the outset to attract the viewers interest, q.v., [11]. The difference between sufficient and transparent is the fact that there is no minimum quality requirement for sufficient encryption. Encryption schemes which can do sufficient encryption cannot necessarily ensure a certain quality and are thus unable to provide transparent encryption.

Given these different application scenarios it is clear that depending on the goal, a security metric has to fulfill different roles. For example, under the assumption of sufficient encryption, a given security metric would have to evaluate which quality is low enough to prevent a pleasant viewing experience. In contrast, for the transparent encryption case, a metric not only has to assess whether the quality of an image or video is low enough, but also whether the quality is high enough to be useful to attract interest. When it comes to content confidentiality the question of quality is no longer applicable. Content confidentiality requires that image content must not be identified by human or automated recognition. This requirement also has to be maintained for any part of the image. Image metrics, in general, do not deal with such questions but rate the overall image quality, the question of intelligibility is usually not covered at all. A drastic example would be an image where only a small part of the image is partly visible. Classical metrics would judge the whole image and consequently would attribute a high security, even though a part of the image is still recognizable which contradicts content confidentiality. Still, it has to be pointed out that content confidentiality can have different forms. To prevent a face recognition scheme from working properly it is sufficient to protect any facial information in a surveillance video, while humans could still be identified in such a video by using gait recognition. Furthermore, if the appearance of a person has to be concealed entirely, a much stronger extent of protection (i.e., higher security) is required. Finally, confidential encryption cannot be solely assessed with security metrics since the scope goes beyond assessing security based on the visual appearance only. Furthermore, we should note that the application of security metrics on video is performed at a frame by frame basis in the literature. We will adopt this model but should note that for the discussion of confidential encryption motion data is of importance, e.g., in [12] it was shown that a replacement attack combined with motion information can reveal the content of a scene even though the visual content of every frame is encrypted.

Consequently, depending on a given application scenario different properties are required from a security metric and different approaches to construct such a metric might perform better or worse for some applications scenarios. This dependence on the evaluation goal of a security metric is hardly ever discussed in the papers introducing a metric. Sufficient and transparent encryption scenarios have a clear and distinct link to the traditional notion of (low) visual quality, while it is highly questionable or at least doubtful if content confidentiality can be assessed by the classical quality notion. While the lack of relation to spatial areas of most security metrics could be compensated in the design to provide locally varying results, the lack of relation to intelligibility in general can probably not be easily resolved.

For both, security metrics and regular image metrics, in the literature we do not find any evaluation whether a given metric can perform the claimed function or how such an evaluation correlates to actual security. However, for regular image metrics it is well known that the correlation with human observations over the full range of possible quality (from high to low quality) does not imply a good performance on a given subset. More specifically, it was pointed out recently that most image metrics perform very poorly for the low quality range ([13]–using the low quality end of the LIVE database). For security metrics, not even this question has been covered so far.

In this paper, we will try to remedy this situation by giving an overview of requirements regarding security goals and formulating these requirements into a testing methodology. Based on this methodology we will evaluate the various security metrics in the literature as well as applicable conventional image metrics. However, we will not deal with every application scenario equally explicitly. We will only make a first step to cover the content confidentiality scenario. The main reason for this is a lack of ground truth. It is not obvious how to generate ground truth for this scenario since there is a disparity between how an image metric works and what is necessary to evaluate content confidentiality. Image metrics, and as an extension security metrics, measure the quality of an image respective to human judgement. This works well for high quality images but suffers for low quality images where human observers can have difficulties differentiating between the severity of an impairment. Thus the methodology to systematically generate ground truth based on human observation needs to be changed for content confidentiality which is not in the scope of this paper. On the other hand, for the image quality-related scenarios (sufficient and transparent encryption), ground truth data is available, in the form of image impairment databases with mean opinion scores (MOS) based on a number of human observations.

In the following we will motivate and introduce a methodology