

On Authentication With Distortion Constraints

Emin Martinian¹, Brian Chen, and Gregory W. Wornell
 Dept. EECS, MIT, Cambridge, MA 02139
 Email: {emin, bchen, gww}@allegro.mit.edu

Abstract — In many multimedia applications, there is a need to authenticate sources subjected to benign degradations such as noise, compression, etc., in addition to potential tampering attacks. Authentication can be enabled through the embedding of suitably chosen markings in the original signal. We develop one information-theoretic formulation of this problem, and identify and interpret the associated fundamental performance limits.

I. INTRODUCTION

As a motivating example, consider the authentication of drivers' licenses. Many jurisdictions print a hologram on the photograph portion of the license. The presence of the hologram indicates that the license is legitimate but does not add excessive distortion. Imprinting a hologram on a license is a particular implementation of a larger class of authentication schemes.

More generally, special markings are embedded into the photograph. A decoder uses these markings to extract an authentic representation of the original. The special markings should be embedded so that the distortion between the original and embedded photographs is small; thus, someone without the appropriate decoder can still use the license to check the identity of the bearer. In addition, the special markings need to be robust to perturbations in the form of smudges or other degradation due to routine handling: the decoder should still declare the photo authentic if only these are present. Finally the special markings should be inserted so that an unauthorized agent can not create a successful forgery.

The idea of inserting special markings in a signal is also used in digital watermarking for copy-protection and data hiding [3]. However, two distinguishing features of the authentication problem are: 1) the attacker's goal is to substantially change the signal in such a way that the decoder is fooled into declaring the result authentic; and 2) whether the attacker can detect the markings added by the encoder is not directly relevant provided that the attacker can not create a convincing forgery.

II. PROBLEM MODEL AND RESULTS

We model the multimedia authentication problem as shown in Figure 1. The original source, X_1^n , is i.i.d. with a known distribution $p_X(x)$. The encoder modifies X_1^n to produce Y_1^n , which then passes through a noisy channel with a known, memoryless probability distribution $p_{Z|Y}(z|y)$. A malicious attacker may use X_1^n , Y_1^n , and Z_1^n to arbitrarily replace the output of the noisy channel with a forgery. Hence we represent the received signal, W_1^n , as the output of an insecure channel. The source, X_1^n , as well as Y_1^n , Z_1^n , and W_1^n take values in a finite alphabet.

Average distortion is measured according to bounded, single-letter distortion measures $d_1(\cdot, \cdot)$, and $d_2(\cdot, \cdot)$. If no decoder is available, authenticity can not be verified and the relevant embedding distortion is $D_1 = \frac{1}{n} \sum_i d_1(X_i, Z_i)$ ². If available, a decoder can attempt to extract an authentic representation of the original. If the

This work has been supported in part by MIT Lincoln Laboratory, NSF under Grant No. CCR-0073520, Microsoft Research, and an NSF Fellowship.

²Without a probabilistic model or distortion constraint for the attacker, we focus on the average distortion between X_1^n and W_1^n in the case where no tampering occurs and $W_1^n = Z_1^n$.

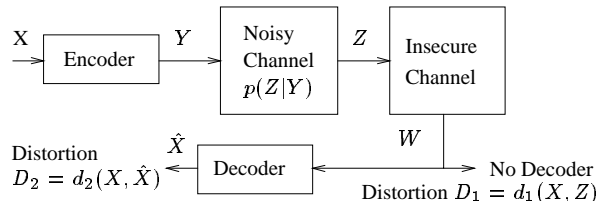


Fig. 1: A diagram of the authentication problem.

decoder determines \hat{X}_1^n to be authentic then the relevant distortion is $D_2 = \frac{1}{n} \sum_i d_2(X_i, \hat{X}_i)$, where clearly for good decoders $D_2 \leq D_1$. If the tampering due to the attacker or the noise is too severe, the decoder might not be able to extract an authentic representation of the original. When this occurs, the decoder outputs a special symbol, \emptyset , to indicate that the received signal is potentially a forgery.

Performance is measured according to three criteria: security, robustness, and distortion. For a scheme to be secure, it should be impossible or infeasible for an attacker to trick the decoder into accepting a forged value for \hat{X}_1^n . For a scheme to be robust, the decoder should almost never declare a signal to be a forgery unless the attacker has tampered with the signal. Finally, a good scheme should keep the distortion between the original source and the received signal as small as possible.

Our main result is an information-theoretic characterization of the tradeoffs between the three criteria. Specifically, authentication schemes that are secure, robust and satisfy a distortion constraint pair (D_1, D_2) are asymptotically achievable if and only if there exists a test channel $p_{Y|X}(y|x)$ and a scalar function $f(\cdot)$ such that

$$\begin{aligned} I(X; Y) &\leq I(Y; Z), \\ E[d_1(X, Z)] &\leq D_1, \\ E[d_2(X, f(Y))] &\leq D_2. \end{aligned}$$

The scalar function $f(\cdot)$ corresponds to a second stage of a two-stage decoder whose first stage recovers Y_1^n from W_1^n , and whose second stage estimates X_1^n from Y_1^n . For example, if the distortion metric was mean square error, then $f(\cdot)$ would correspond to the scalar minimum mean-square estimator of X given Y .

Practically, this analysis underscores the important role that coding has to play in designing authentication systems that approach the optimal tradeoffs. Specifically, in [1] we show that relatively simple trellis codes can obtain more than 5 dB less distortion than a comparable uncoded scheme with the same levels of security and robustness.

REFERENCES

- [1] E. Martinian, B. Chen, and G. W. Wornell, "Information Theoretic Approach To The Authentication Of Multimedia," in *Proc. of SPIE: EI-01*, San Jose, CA, Jan. 2001.
- [2] E. Martinian, *Authenticating Multimedia In The Presence Of Noise*. SM thesis, MIT, Cambridge, MA June 2000.
- [3] B. Chen and G. W. Wornell, "Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding," *IEEE Trans. Inform. Theory*, May 2001.