IMAGE AUTHENTICATION BASED ON DISTRIBUTED SOURCE CODING

Yao-Chung Lin, David Varodayan, and Bernd Girod

Information System Laboratory, Stanford University, Stanford, CA 94305
{yao-chung.lin, varodayan, bgirod}@stanford.edu

ABSTRACT

Image authentication is important in content delivery via untrusted intermediaries, such as peer-to-peer (P2P) file sharing. Many differently encoded versions of the original image might exist. On the other hand, intermediaries might tamper with the contents. Distinguishing the legitimate diversity of encodings from malicious manipulation is the challenge addressed in this paper.

We develop a novel approach based on distributed source coding for the problem of backward-compatible image authentication. The key idea is to provide a Slepian-Wolf encoded quantized image projection as authentication data. This version can be correctly decoded only with the help of an authentic image as side information. Distributed source coding provides the desired robustness against legitimate encoding variations, while detecting illegitimate modification. We demonstrate false acceptance rates close to zero for authentication data sizes that are only a few percent of the compressed image size.

Index Terms— Image authentication, distributed source coding, LDPC codes

1. INTRODUCTION

Media authentication is important in many applications of content delivery via untrusted intermediaries, such as peer-to-peer (P2P) file sharing or P2P multicast streaming. In these applications, many differently encoded versions of the original media file might exist. Moreover, transcoding and bitstream truncation at intermediate nodes might be required, giving rise to further diversity. On the other hand, intermediaries might tamper with the contents for a variety of reasons, such as interfering with the distribution of a particular file, piggybacking unauthentic content, or generally discrediting a particular distribution system. Distinguishing the legitimate diversity of encodings from malicious manipulation is the major technical challenge for media authentication systems. Past approaches fall into two groups: watermarks and media hashes.

This work has been supported, in part, by a gift from NXP Semiconductors to the Stanford Center for Integrated Systems.

A "fragile" watermark can be embedded into the host signal waveform without perceptual distortion [1] [2]. Users can confirm the authenticity by extracting the watermark from the received content. The system design should ensure that the watermark survives lossy compression, but that it "breaks" as a result of a malicious manipulation. Unfortunately, watermarking authentication is not backward compatible with previously encoded contents; unmarked contents cannot be authenticated later. Embedded watermarks might also increase the bit-rate required when compressing a media file.

Media hashing [3] [4] achieves verification of previously encoded media by using an authentication server to supply authentication data to the user. Media hashes are inspired by cryptographic digital signatures [5], but unlike cryptographic hash functions, media hash functions are supposed to offer proof of perceptual integrity. Using a cryptographic hash, a single bit difference leads to an entirely different hash value. If two media signals are perceptually indistinguishable, they should have identical hash values. A common approach of media hashing is extracting features which have perceptual importance and should survive compression. The authentication data are generated by compressing these features or generating their hash values. The user checks the authenticity of the received content by comparing the features or their hash values to the authentication data.

We propose an extension of hashing for image authentication based on distributed source coding. The system has similarities with the secure biometric authentication scheme in [6]. It is also related to the semi-fragile watermarking scheme for images in [7], which, however, is not applicable to authentication of previously encoded images. In our proposal, the authentication server provides a user with a Slepian-Wolf encoded image projection, and the user attempts to decode this bitstream using the image-to-be-authenticated as side information. The Slepian-Wolf result [8] indicates that the lower the distortion between side information and the original, the fewer authentication bits are required for correct decoding. By correctly choosing the size of the authentication data, this insight allows us to distinguish between legitimate encoding variations of the image and illegitimate modifications. In Section 2, we will describe the proposed image authentication scheme and its rationale in detail. Simulation results will be presented in Section 3.

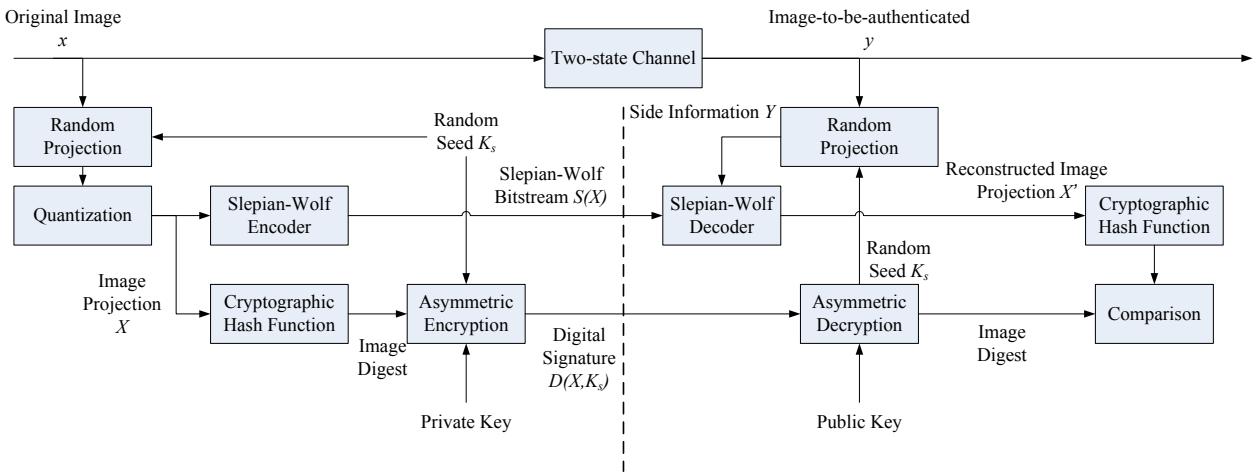


Fig. 1. Image authentication system based on distributed source coding

2. PROPOSED IMAGE AUTHENTICATION SCHEME

Fig. 1 depicts our proposed image authentication scheme. We denote the source image as x. The user receives the image-to-be-authenticated y as the output of a two-state lossy channel that models legitimate and illegitimate modifications. The left-hand side of Fig. 1 shows that the authentication data consist of a Slepian-Wolf encoded quantized image projection of x and a digital signature of that version. The verification decoder, in the right-hand side of Fig. 1, knows the statistics of the worst permissible legitimate channel and can correctly decode the authentication data only with the help of an authentic image y as side information.

2.1. Two-State Channel

We model the image-to-be-authenticated y by way of a two-state lossy channel, shown in Fig. 2. In the legitimate state, the channel performs lossy compression and reconstruction, such as JPEG and JPEG2000, with peak signal-to-noise ratio (PSNR) of 30 dB or better. In the illegitimate state, it additionally includes a malicious attack.

Fig. 3 compares a sample input and two outputs of this channel. The source image x is "Lena" at 512x512 resolution. In the legitimate state, the channel is JPEG2000 compression and reconstruction at (the worst permissible) 30dB PSNR. In the illegitimate state, a further malicious attack is applied: a 32x100 pixel text banner is overlaid on the reconstructed image.

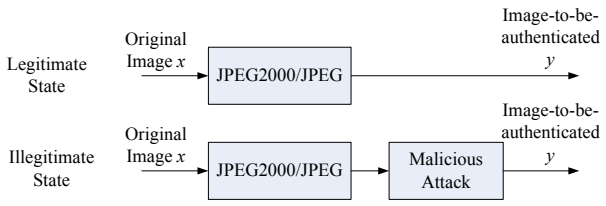
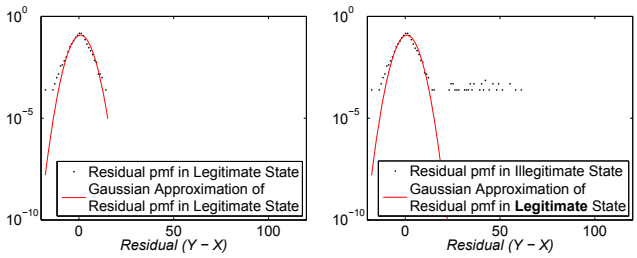


Fig. 2. Two-state lossy channel

The joint statistics of x and y vary depending on the state of the channel. We illustrate this by plotting in Fig. 4 the distribution of the residual $D = Y - X$, where X and Y are image projections of x and y in Fig. 3, respectively. The projection is a blockwise pseudorandomly weighted mean and will be described in detail in the next section. Since the legitimate channel consists of JPEG2000 or JPEG compression and reconstruction, the samples of the projection residual D are weighted sums of quantization errors. Therefore, the distribution of D resembles a Gaussian, by the central limit theorem. In the illegitimate channel state, the image samples in the tampered region are unrelated to those of the original image, giving the distribution of D non-negligible tails. It is the modification of the joint statistics of X and Y that is exploited for authentication.

(a) (b) (c)

Fig. 3. Portion of "Lena" image (a) x original, (b) y at output of legitimate channel, (c) y at output of illegitimate channel



(a) (b)

Fig. 4. The residual distributions between downsampled channel input and output, (a) in legitimate state, (b) in illegitimate state

2.2. Authentication Data Generation and Verification

In our authentication system shown in Fig. 1, a pseudorandom projection (based on a randomly drawn seed K_s) is applied to the original image x and the projection coefficients are quantized to yield X. The authentication data comprise two parts, both derived from X. The Slepian-Wolf bitstream $S(X)$ is the output of a Slepian-Wolf encoder based on low-density parity-check (LDPC) codes [9] and the much smaller digital signature $D(X, K_s)$ consists of the seed K_s and a cryptographic hash value of X signed with a private key.

The authentication data are generated by a server upon request. Each response uses a different random seed K_s, which is provided to the decoder as part of the authentication data. This prevents an attack which simply confines the tampering to the nullspace of the projection. Based on the random seed, for each 16x16 nonoverlapping block B_i, we generate a 16x16 pseudorandom matrix P_i by drawing its elements independently from a Gaussian distribution $\mathcal{N}(1, \sigma_z^2)$ and normalizing so that $||P_i||_2 = 1$. We choose $\sigma_z = 0.2$ empirically. The inner product $\langle B_i, P_i \rangle$ is quantized into an element of X.

The rate of the Slepian-Wolf bitstream $S(X)$ determines how statistically similar the image-to-be-authenticated must be to the original to be declared authentic. If the conditional entropy $H(X|Y)$ exceeds the bit-rate R in bits per pixels, X can no longer be decoded correctly [8]. Therefore, the rate of $S(X)$ should be chosen to distinguish between the different joint statistics induced in the images by the legitimate and illegitimate channel states. At the encoder, we select a Slepian-Wolf bit-rate just sufficient to authenticate both legitimate 30dB JPEG2000 and JPEG reconstructed versions of the original image.

At the receiver, the user seeks to authenticate the image y with authentication data $S(X)$ and $D(X, K_s)$. It first projects y to Y in the same way as during authentication data generation. A Slepian-Wolf decoder reconstructs X' from the Slepian-Wolf bitstream $S(X)$ using Y as side information. Decoding is via LDPC belief propagation [9] initialized according to the statistics of the legitimate channel state at the worst permissible quality for the given original image. Finally, the image digest of X' is computed and compared to the image digest, decrypted from the digital signature $D(X, K_s)$ using a public key. If these two image digests are identical, the receiver recognizes image y as authentic.

3. SIMULATION RESULTS

We use the test images "Barbara", "Lena", "Mandrill", and "Peppers" at 512x512 resolution in 8-bit gray resolution. The two-state channel in Fig. 2 has JPEG2000 or JPEG compression and reconstruction applied at several qualities. The malicious attack consists of the overlaying of a 32x100 text banner at a random location in the image. The text color is white or black, depending on which is more visible. This avoids generating trivial attacks, such as overlaying a white text on a white area. The encoder quantization is varied so that the Slepian-Wolf encoder processes between 1 to 8 bitplanes, starting with the most significant. The Slepian-Wolf codec is implemented using LDPC Accumulate (LDPCA) codes [10] with block size of 1024 bits. During authentication data generation, the bitplanes of X are encoded successively as LDPCA syndromes. The bitplanes are conditionally decoded, each decoded bitplane acting as additional side information for subsequent bitplanes, as in [11].

Fig. 5 compares the minimum rate that would be required to decode the Slepian-Wolf bitstream $S(X)$ for side information Y due to legitimate and illegitimate channel states, for the image "Lena" quantized to 3 bitplanes. The following observations also hold for other images and levels of quantization. The rate required to decode $S(X)$ with legitimately created side information is significantly lower than the rate (averaged over 100 trials) when the side information is illegitimate, for JPEG2000 or JPEG reconstruction PSNR above 30dB. Moreover, as the PSNR increases, the rate for legitimate side information decreases, while the rate for illegitimate side information stays high. The rate gap justifies our choice for the Slepian-Wolf bitstream size: the size just sufficient to authenticate both legitimate 30dB JPEG2000 and JPEG reconstructed versions of the original image.

Fig. 6 shows this selected Slepian-Wolf bitstream size in bytes for the four test images at quantization levels ranging from 1 to 8 bitplanes. For 4 bitplanes, the Slepian-Wolf bitstream size is less than 66 bytes or 2.3% of the encoded file sizes at 30dB reconstruction.

We now measure the false acceptance rate of our authentication system; that is, the proportion of illegitimately modified images declared to be authentic. Table 1 shows results for each test image at each quantization level (from 1 to 4 bitplanes) using 3000 trials of varying reconstruction quality (at least 30dB), compression method (JPEG or JPEG2000), and location for the overlaid text. For all images, 4 bitplane quantization is sufficient to reduce the false acceptance rate to zero; this corresponds to a Slepian-Wolf bitstream size of less than

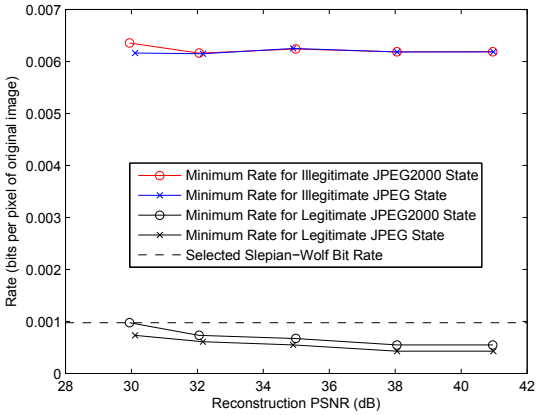


Fig. 5. Minimum rate for decoding Slepian-Wolf bitstream for the image "Lena" quantized to 3 bitplanes

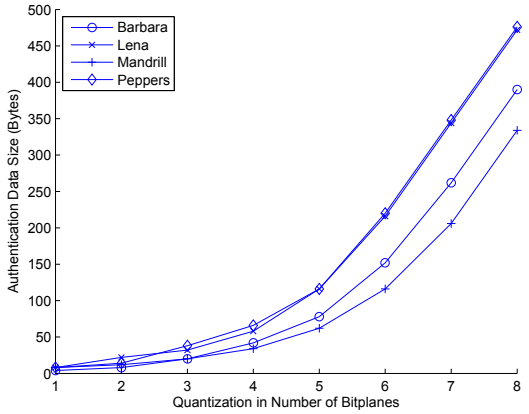


Fig. 6. Slepian-Wolf bitstream size in bytes

66 bytes (see Fig. 6) or 2.3% of the compressed image size. The proportion of authentic images declared to be inauthentic is the false rejection rate; it is zero as long as the compression reconstruction quality of the image-to-be-authenticated is at least 30dB, due to our choice of Slepian-Wolf bitstream size.

4. CONCLUSIONS

In this work, we developed a novel backward-compatible image authentication scheme, based on distributed source coding, that distinguishes between legitimate encoding variations of an image and illegitimately modified versions. We demonstrated false acceptance rates close to zero for authentication data size less than 66 bytes or 2.3% of the compressed image size. We intend to extend this scheme to authentication of video sequences in P2P settings.

Table 1. False acceptance rate for test images at different quantizations

	Quantization in number of bitplanes			
	1	2	3	4
Barbara	6.23%	0.57%	0.00%	0.00%
Lena	4.10%	1.90%	0.10%	0.00%
Mandrill	0.73%	0.20%	0.00%	0.00%
Peppers	5.87%	0.97%	0.00%	0.00%

5. REFERENCES

[1] J.J. Eggers and B. Girod, "Blind watermarking applied to image authentication," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, UT, May 2001.

[2] R. B. Wolfgang and E. J. Delp, "A watermark for digital images," in *IEEE International Conference on Image Processing*, Lausanne, Switzerland, Sep. 1996.

[3] C.-Y. Lin and S.-F. Chang, "A robust image authentication method distinguishing JPEG compression from malicious manipulation," *IEEE Transcations on Circuits and Systems for Video Technology*, vol. 11, no. 2, pp. 153–168, Feb. 2001.

[4] C.-S. Liu and H.-Y. M. Liao, "Structural digital signature for image authentication: an incidental distortion resistant scheme," *IEEE Transcations on Multimedia*, vol. 5, no. 2, pp. 161–173, June 2003.

[5] W. Diffie and M. E. Hellman, "New directions in cryptography," *IEEE Transcations on Information Theory*, vol. IT-22, no. 6, pp. 644–654, Jan. 1976.

[6] E. Martinian, S. Yekhanin, and J. S. Yedidia, "Secure biometrics via syndromes," in *Allerton Conference on Communications, Control and Computing*, Monticello, IL, Sep. 2005.

[7] Q. Sun, S.-F. Chang, M. Kurato, and M. Suto, "A new semi-fragile image authentication framework combining ECC and PKI infrastructure," in *IEEE International Symposium on Circuits and Systems*, Phoenix, AZ, May 2002.

[8] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. IT-19, no. 4, pp. 471–480, July 1973.

[9] A. Liveris, Z. Xiong, and C. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Communications Letters*, vol. 6, no. 10, pp. 440–442, Oct. 2002.

[10] D. Varodayan, A. Aaron, and B. Girod, "Rate-adaptive codes for distributed source coding," *EURASIP Signal Processing Journal, Special Section on Distributed Source Coding*, vol. 86, no. 11, pp. 3123–3130, Nov. 2006.

[11] A. Aaron, S. Rane, E. Setton, and B. Girod, "Transform-domain Wyner-Ziv codec for video," in *SPIE Visual Communications and Image Processing Conference*, San Jose, CA, 2004.