DISTRIBUTED SOURCE CODING AUTHENTICATION OF IMAGES WITH CONTRAST AND BRIGHTNESS ADJUSTMENT AND AFFINE WARPING

Yao-Chung Lin, David Varodayan, and Bernd Girod

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

ABSTRACT

Media authentication is important in content delivery via untrusted intermediaries, such as peer-to-peer (P2P) file sharing. Many differently encoded versions of a media file might exist. Our previous work applied distributed source coding not only to distinguish the legitimate diversity of encoded images from tampering but also to localize tampered regions in an image already deemed to be inauthentic. In both cases, authentication requires a Slepian-Wolf encoded image projection that is supplied to the decoder.

We extend our scheme to authenticate images that have undergone contrast, brightness, and affine warping adjustment. Our approach incorporates an Expectation Maximization algorithm into the Slepian-Wolf decoder. Experimental results demonstrate that the proposed algorithm can distinguish legitimate encodings of authentic images from illegitimately modified versions, despite arbitrary contrast, brightness, and affine warping adjustment, using authentication data of less than 250 bytes per image.

Index Terms— Image authentication, distributed source coding, Expectation Maximization

1. INTRODUCTION

Media authentication is important in 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 file might exist. Moreover, transcoding and bitstream truncation at intermediate nodes might give rise to further diversity. Intermediaries might also tamper with the media for a variety of reasons, such as interfering with the distribution of a particular file, piggybacking unauthentic content, or generally discrediting a legitimacy of reasons, such as interfering with the distribution of a particular file, piggybacking unauthentic content, or generally discrediting a

2. REVIEW OF IMAGE AUTHENTICATION WITH DSC

Fig. 1 is the block diagram for our earlier image authentication scheme [1] as well as the current work. We denote the source image as $x$. We model the target image $y$ by way of the space-varying two-state lossy channel in Fig. 2. The legitimate state of the channel performs lossy JPEG2000 or JPEG compression and reconstruction with peak signal-to-noise ratio (PSNR) of 30 dB or better. The tampered state additionally includes malicious tampering. The channel state variable $S_i$ is defined per nonoverlapping 16x16 block of image $y$. If any pixel in block $B_i$ has been tampered with, $S_i = 1$; otherwise, $S_i = 0$.

We now review the authentication system. The left-hand side of Fig. 1 shows that a pseudorandom projection (based on a randomly drawn seed $K_S$) is applied to the original image $x$ to produce projection coefficients $X$, which are quantized to $X_q$. The authentication data comprise two parts, both derived from $X_q$. The Slepian-Wolf bitstream $S(X_q)$ is the output of a Slepian-Wolf encoder based on rate-adaptive low-density parity-check (LDPC) codes [6]. The much smaller digital signature $D(X_q, K_s)$ consists of the seed $K_s$ and a cryptographic hash value of $X_q$ 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 $\mathcal{N}(1, \sigma^2)$ and normalizing so that $\|P_i\|_2 = 1$. We choose $\sigma = 0.2$ empirically. The inner product $\langle B_i, P_i \rangle$ is an element of $X$, quantized to an element of $X_q$.

The authentication decoder, on the right-hand side of Fig. 1, seeks to authenticate the image $y$ with authentication data $S(X_q)$ and $D(X_q, K_s)$. It first projects $y$ to $Y$ in the same way as during authentication data generation. A Slepian-Wolf decoder reconstructs $X_q'$ from the Slepian-Wolf bitstream $S(X_q)$ using $Y$ as side information. Decoding is via joint bitplane LDPC belief propagation [7] initialized according to the known statistics of the legitimate channel state at the worst permissible quality for the given original image. Then the image digest of $X_q'$ is computed and compared to the image digest, decrypted from the digital signature $D(X_q, K_s)$ using a public key. If these two image digests are not identical, the receiver declares image $y$ to be inauthentic. If they match, then $X_q$ has been recovered. To confirm the authenticity of $y$, the receiver verifies that the likelihood of the legitimate channel is greater than a certain threshold.

Since this second-pass comparison uses all available information, the threshold for the likelihood specifies how statistically similar the target image must be to the original to be declared...
authentic. But the rate of the Slepian-Wolf bitstream $S(X_q)$ determines whether the quantized image projection $X_q$ is recovered at all [8]. Accordingly, at the encoder, we select a Slepian-Wolf bitrate just sufficient to successfully decode with both legitimate 30 dB JPEG2000 and JPEG reconstructed versions of $x$. At the decoder, we choose a threshold for the likelihood for the second-pass comparison to distinguish between the different joint statistics induced in the images by the legitimate and tampered channel states.

3. CONTRAST, BRIGHTNESS, AND AFFINE WARPING ADJUSTMENT MODEL

In this paper, we replace the two-state lossy channel in Fig. 2 with the one in Fig. 3. Now both the legitimate and tampered states of the channel are affected by contrast and brightness adjustment and affine warping. In the legitimate state, we model the channel as

$$y(m) = \alpha x(n) + \beta + z(m),$$

where $m, n \in R^2$, and $\alpha, \beta \in R$, with $n = A m + b$, where $A \in R^{2 \times 2}$, $b \in R^2$.

$A$ and $b$ are transformation and translation parameters in affine warping, respectively, $\alpha$ and $\beta$ are contrast and brightness adjustment parameters, and $z$ is noise introduced by compression and reconstruction. To keep $y$ the same size as $x$, it is padded with black pixels (arbitrarily) and cropped. Fig. 4(a) shows a source image “Lena” at 512x512 original resolution. The legitimate $y$ in Fig. 4(b) is first contrast and brightness adjusted with $(\alpha, \beta) = (1.2, -20)$, rotated by 5 degrees around the image center, then cropped to 512x512, and finally JPEG2000 compressed and reconstructed at 32 dB PSNR. Here, $A = \begin{bmatrix} 0.996 & -0.087 \\ 0.087 & 0.996 \end{bmatrix}$ and $b = [23 \ -21]^T$. The tampered $y$ in Fig. 4(c) additionally includes malicious tampering. Fig. 4(d) shows the tampered $y$ realigned to the original, with channel states $S_i$ labeled red if tampered and blue if cropped out in $y$. The rest are legitimate cropped-in states.

The image authentication system described in Section 2 cannot authenticate legitimate images that have undergone contrast, brightness, and affine warping adjustment, because the side information is neither aligned with the corresponding authentication data nor compensated for the contrast and brightness changes. Approaches suggested in the prior art to overcome this problem involve generating affine-invariant features that serve as authentication data [9, 10, 11]. These features are invariant to some transformations, but would be sensitive to others. We instead propose that the authentication decoder jointly estimate the adjustment parameters directly from the Slepian-Wolf bitstream $S(X_q)$ and the target image $y$ using an EM algorithm.

4. EXPECTATION MAXIMIZATION

The introduction of learning to the system in Fig. 1 requires a modification of the Slepian-Wolf bitstream block from a joint-bitplane LDPC decoder [7] to the parameter-learning Slepian-Wolf decoder shown in Fig. 5. As before, it takes the Slepian-Wolf bitstream $S(X_q)$ and the target image $y$ and yields the reconstructed image projection $X_q'$. But it now also estimates the parameters via an EM algorithm. The E-step updates the $a \ posteriori$ probability mass functions (pmf) $P_{app}(X_q)$ using the joint bitplane decoder and also estimates corresponding coordinates for a subset of reliably-decoded projection pixels. The M-step updates the parameters based on the corresponding coordinate pmfs, denoted $Q(m)$ in Fig. 5. This loop of EM iterations terminates when hard decisions on $P_{app}(X_q)$ satisfy the constraints imposed by $S(X_q)$.
Fig. 4. Test image “Lena” (a) x original, (b) y in legitimate state, (c) y in tampered state, (d) channel states S_i (red: tampered, blue: cropped-out) associated with the 16x16 blocks of realigned output in (c).

In the E-step, we fix the parameters $A, b, \alpha, \beta$ at their current hard estimates. The inverse adjustment is applied to the image $y$ to obtain a compensated image $y_{\text{comp}}$. If the parameters are accurate, $y_{\text{comp}}$ would be closely aligned to the original image $x$ in the cropped-in region. We derive intrinsic pmfs for the image projection pixels $x_q$ as follows. In the cropped-in region, we use Gaussian distributions centered at the random projection values of $y_{\text{comp}}$, and in the cropped-out region, we use uniform distributions. Then, we run three iterations of joint bitplane LDPC decoding on the intrinsic pmfs $P_{\text{app}}([X_q]_i = x_q)$.

We estimate the corresponding coordinates $m^{(i)}$ for those projection pixels for which $\max_{x_q} P_{\text{app}}([X_q]_i = x_q) > T = 0.995$, denoting this set of reliably-decoded projection indices as $\mathcal{C}$. We also denote the maximizing reconstruction value $x_q$ to be $[x_q^{\max}]$. The latent variable update can be written as $Q_i(m) := P(m^{(i)} = m|y^{(i)}; A, b, \alpha, \beta)$, where $y^{(i)}$ is the set of top-left coordinates of the 16x16 projection blocks $B_i$ in the original image $x$, and $m^{(i)}$ represents the corresponding set of coordinates in the overcomplete random projection $Y$.

In the M-step, we re-estimate the parameters $A, b, \alpha, \beta$ by holding the corresponding coordinate pmfs $Q_i(m)$ fixed and maximizing a lower bound of the log-likelihood function:

$$L(A, b, \alpha, \beta) \equiv \sum_{i \in \mathcal{C}} \log P([x_q^{\max}]_i, y^{(i)}; A, b, \alpha, \beta)$$

$$= \sum_{i \in \mathcal{C}} \log \left( \sum_{m^{(i)}} P([x_q^{\max}]_i, y^{(i)}|m^{(i)}; A, b, \alpha, \beta)P(m^{(i)}) \right)$$

$$\geq \sum_{i \in \mathcal{C}} \sum_{m^{(i)}} Q_i(m) \log P([x_q^{\max}]_i, y^{(i)}|m^{(i)}; A, b, \alpha, \beta)$$

$$= \sum_{i \in \mathcal{C}} \sum_{m^{(i)}} Q_i(m) \left( \log P(n^{(i)}|m^{(i)}; y_q^{(i)}; A, b) + \log P([x_q^{\max}]_i, y|m; \alpha, \beta) \right).$$

The lower bound is due to Jensen’s inequality and concavity of log(). Note also that $P([x_q^{\max}]_i, y|m^{(i)}; \alpha, \beta)$ does not depend on the parameters $A$ and $b$, and $P(n^{(i)}|m, x_q^{\max}]; y; A, b)$ does not depend on the parameters $\alpha$ and $\beta$. Thus, we can maximize

\begin{align*}
\alpha &:= \frac{1}{|\mathcal{C}|} \sum_{i \in \mathcal{C}} \mu^{(i)}_X - \frac{1}{|\mathcal{C}|} \sum_{i \in \mathcal{C}} \sum_{j \in \mathcal{C}} \mu^{(i)}_X \mu^{(j)}_Y \\
\beta &:= \frac{1}{|\mathcal{C}|} \sum_{i \in \mathcal{C}} \mu^{(i)}_Y - \alpha \mu^{(i)}_X,
\end{align*}

where $\mu^{(i)}_X := E_m-Q_i[E[X|Y(m), [x_q^{\max}]_i]]$, $\mu^{(i)}_Y := E_m-Q_i[E[Y(m), [x_q^{\max}]_i]]$, $\mu^{(i)}_X := E_m-Q_i[E[X|Y(m), [x_q^{\max}]_i]]$, $\mu^{(i)}_Y := E_m-Q_i[E[X|Y(m), [x_q^{\max}]_i]]$. 

Fig. 5. Slepian-Wolf decoder with parameter learning.
5. SIMULATION RESULTS

Our experiments use “Barbara”, “Lena”, “Mandrill”, and “Peppers” of size 512x512 at 8-bit gray resolution. The two-state channel in Fig. 3 changes contrast and brightness, applies affine warping to the images, and crops them to 512x512. Then JPEG2000 or JPEG compression and reconstruction is applied at 30 dB reconstruction PSNR. In the tampered state, the malicious attack overlays a 20x122 pixel text banner, white or black whichever is more visible, randomly on the image. The image projection $X$ is quantized to 4 bits, and the Slepian-Wolf encoder uses a 4096-bit LDPC code with 740 degree-1 syndrome nodes.

Fig. 6 compares the minimum rates for decoding $S(X_q)$ with legitimate test images using three different decoding schemes: the proposed EM decoder that learns all the parameters in contrast and brightness adjustment and affine warping, an oracle decoder that knows the parameters, and a fixed decoder that always assumes no change applied in the images except compression. Fig. 6 (a) and (b) show the results when the affine warping is rotation around the image center and horizontal shearing, respectively. In both cases, contrast is increased by 20%, and brightness is decreased by 20 out of 255. The EM decoder requires minimum rates only slightly higher than the oracle decoder, while the fixed decoder requires higher rates as the geometric distortion increases and quickly reaches a regime, where the side information does no longer reduce the bit-rate of the quantized random projections (0.015625 bits per pixel of the original image).

![Graph showing the minimum rates for decoding S(X_q) with legitimate test images using different decoding schemes.](image)

Fig. 6. Minimum rate for decoding the legitimate test image, “Barbara,” using different decoders.

For the next experiment, we set the authentication data size to 245 bytes and measure false acceptance and rejection rates. The acceptance decision is made based on the likelihood of $X_q$ and $Y$ with estimated parameters within the estimated cropped-in blocks. The channel settings remain the same except that parameters $\alpha$ is randomly drawn from [0.9, 1.1], $\beta$ from [−10, 10], $A_{11}$ and $A_{22}$ from [0.95, 1.05], $A_{21}$ and $A_{12}$ from [−0.05, 0.05], and $b_1$ and $b_2$ from [−10, 10]. The JPEG2000/JPEG reconstruction PSNR is selected from 30 to 42 dB. With 2000 trials each on “Barbara”, “Lena”, “Mandrill”, and “Peppers,” Fig. 7 shows the receiver operating characteristic curves created by sweeping the decision threshold of the legitimate likelihood. The EM decoder performs very closely to the oracle decoder, while the fixed decoder rejects authentic test images with high probability. In the legitimate case, the EM decoder estimates the transform parameters $A_{11}$, $A_{21}$, $A_{12}$, $A_{22}$, $b_1$, $b_2$, $\alpha$, and $\beta$ with mean squared error $4.5 \times 10^{-7}$, $2.6 \times 10^{-6}$, $3.4 \times 10^{-7}$, $1.6 \times 10^{-6}$, 0.05, 0.54, $2 \times 10^{-5}$, 0.34 and respectively.

![Graph showing receiver operating characteristic curves.](image)

Fig. 7. Receiver operating characteristic curves.

6. CONCLUSIONS

We have extended our image authentication system to handle images that have undergone both contrast and brightness adjustment and affine warping. Our authentication decoder learns the contrast and brightness adjustment and affine warping parameters via an unsupervised EM algorithm. We demonstrate that an authentication Slepian-Wolf bitstream of 245 bytes is sufficient to distinguish between legitimate editings of images and illegitimately modified versions. The work can be extended to other editing models using an appropriate M-Step.

7. REFERENCES


