June 2004

An MPEG tolerant authentication system for video data

Takeyuki Uehara
University of Wollongong, takeyuki@uow.edu.au

R. Safavi-Naini
University of Wollongong, rei@uow.edu.au

P. Ogunbona
University of Wollongong, philipo@uow.edu.au

An MPEG tolerant authentication system for video data

Abstract

We propose a secure video authentication algorithm that is tolerant to visual degradation due to MPEG lossy compression to a designed level. The authentication process generates a tag that is sent with video data and the level of protection can be adjusted so that longer tags are used for higher security, and that the protection is distributed such that higher security is provided for regions of interest in the image. The computation required for authentication and verification can be largely performed as part of MPEG compression and so generation and verification of the tag can be integrated into the compression system. Calculation of the tag can be parallelized and so made fast.

Keywords

data compression digital signatures, parallel processing, video coding, watermarking

Disciplines

Physical Sciences and Mathematics

Publication Details

This paper appeared as: Uehara, T, Safavi-Naini, R & Ogunbona, P, An MPEG tolerant authentication system for video data, IEEE International Conference on Multimedia and Expo, 27-30 June 2004, 2, 891-894.

An MPEG Tolerant Authentication System for Video Data

Takeyuki Uehara, Reihaneh Safavi-Naini and Philip Ogunbona
School of Information Technology and Computer Science,
University of Wollongong, Wollongong, NSW 2522, Australia,
email:{takeyuki,rei,philipo}@uow.edu.au

Abstract— **We propose a secure video authentication algorithm that is tolerant to visual degradation due to MPEG lossy compression to a designed level. The authentication process generates a tag that is sent with video data and the level of protection can be adjusted so that longer tags are used for higher security, and that the protection is distributed such that higher security is provided for regions of interest in the image. The computation required for authentication and verification can be largely performed as part of MPEG compression and so generation and verification of the tag can be integrated into the compression system. Calculation of the tag can be parallelized and so made fast.**

I. Introduction

In many applications such as news reporting and surveillance, the visual data must be authenticated. Visual data is communicated in compressed form. A *compression tolerant authentication system* will tolerate changes that are due to lossy compression, while detecting other changes. Cryptographic authentication systems are sensitive to a single bit change in data and so cannot be directly used for compression tolerant authentication. In this paper we consider *MPEG tolerant video authentication systems*.

Compression tolerant video authentication systems can be broadly divided into *feature extraction systems* and *watermarking systems*. In the former the authentication system extracts features (also called *signature*, *digest* or *message authentication code (MAC)*) of the video data that remain invariant through lossy compression to the given quality level. An authenticated video stream consists of a compressed video stream and a feature stream. In watermarking approach, a fragile watermark [1] is embedded in the video data and the watermark must be destroyed when the video is tampered with. Reconciling fragility and compression tolerance is a challenging task. Another disadvantage of the watermarking approach is the degradation of quality due to the embedded noise (i.e. the watermark).

In this paper we propose a MAC system for video data that tolerate MPEG compression to a given quality level. We evaluate the effectiveness of the system by considering attacks and show the best known attack is computationally expensive. Although this does not prove security of the system in general but gives an estimate of the cost against known attacks. The system has a number of attractive properties.

1) The main computation of the system is computing the *Discrete Cosine Transform* (DCT) [2] of image blocks which is part of MPEG compression algorithm and so the proposed system can be effectively integrated into MPEG.
2) The computation is parallelizable and the system can be used for real-time authentication of data.
3) The system provides flexible protection. It allows longer MACs to be used for higher level of protection and supports non-uniform protection; that is, selected parts of an image can be protected to a higher level. This is a useful property for protection of regions of interest in images.

Section II gives an overview of MPEG. In Section III we describe our system and show its properties. Section IV describe the feature codes to construct a message authentication system and in Section V we analyze the security of this system. Section VI concludes the paper.

II. MPEG Compression

A video compression standard *MPEG* [3] is a lossy compression system. In an *MPEG video stream*, the image sequence is encoded as a sequence of *intra*, *forward predicted*, and *bidirectional prediction* frames [4]. An *intra frame (I-frame)* is encoded without reference to any other frames; a *forward predicted frame (P-frame)* is encoded relative to the past reference I- or P- frame and a *bidirectional prediction frame (B-frame)* is encoded relative to the past and/or future reference I- or P- frames.

A frame is divided into 16×16 *macroblocks*. A macroblock consists of 4 luminance, and 2 chrominance 8×8 blocks for 4:2:0 chroma format and 4 chrominance blocks for 4:2:2 format. In an I-frame, the transform coding is performed on the macroblocks called *intra macroblocks*. Each of 8×8 blocks in an intra macroblock is transformed into 64 coefficients using DCT. This is followed by a scalar quantizer that replaces each DCT coefficient with an integer. Finally the 64 quantized coefficients are *"zig-zag" scanned* and are entropy-coded. The macroblocks in P- and B- frame are either encoded as a *motion vector* and an *error term* between the macroblock and the area, or intra-coded.

The information loss is primarily due to quantization. However computation error also contributes to the difference between the values of a pixel, before and after the compression.

0-7803-8603-5/04/$20.00 ©2004 IEEE

III. An MPEG Tolerant Authentication Scheme for Video Data

We propose a message authentication code (MAC) that consists of *feature codes* which are obtained by encoding a linear combination of DCT coefficients of subsets of blocks. The MAC tolerates MPEG compression above a given compression quality level.

A. Authentication

8×8 pixel blocks in an I-frame are divided into subsets and DCT coefficients in a subset is used to generate a feature code.

In the following, we assume ${A_i}^{(u)}$ are non-negative integers although the approach can also be used for arbitrary ${A_i}^{(u)}$ values. Let $\{G_1, G_2 \cdots G_{g/m}\}$ be a partition of blocks assuming the subsets have m blocks each. Blocks in a group j are labeled by 1 to m. The feature code generation algorithm is as follows.

1) Find the DCT coefficients of each block.
2) Let ${F_{i,j}}^{(u)}$ denote the DCT coefficient in position (frequency) u of the ith block in G_j. Then ${Y_j}^{(u)} = \sum_{\forall i \in [m]} {A_i}^{(u)} {F_{i,j}}^{(u)}$ is the weighted sum of all coefficients in ${G_j}^{(u)}$.
3) A *feature code* is generated by encoding ${Y_j}^{(u)}$.

The correctness of the message authentication algorithm follows from the observation that the value of a linear sum as defined above, in the original image and its decompressed version, will remain 'close' and this closeness can be estimated. Theorem 1, proved in [5] and re-stated here for completeness, formalizes this statement. To state the theorem we need the following notations.

Denote the set of integers $\{1, 2, 3, ..., m\}$ by $[m]$. Let $\frac{{F_p}^{(u)}}{Q^{(u)}} = R_p = h_p + r_p$ be the DCT coefficient in position u in block p, divided by the corresponding quantization scalar, and $-0.5 \leq r_p < 0.5$.

The original and reconstructed DCT coefficient values, ${F_p}^{(u)}$ and $\tilde{F}_p^{(u)}$, are related as follows.

$$\begin{array}{rcccl} {F_p}^{(u)} & = & R_p Q^{(u)} & = & h_p Q^{(u)} + r_p Q^{(u)} \\ \tilde{F}_p^{(u)} & = & rint(R_p) Q^{(u)} & = & {F_p}^{(u)} - r_p Q^{(u)} \end{array}$$

Let k be a real number. Then $\frac{k}{Q^{(u)}} = \tilde{k}^{(u)} + r^{(u)}$ and $-0.5 \leq r^{(u)} < 0.5$ and the reconstructed value of k is $\tilde{k} = \tilde{k}^{(u)} Q^{(u)}$.

For DC and AC coefficients we have the following two theorems, respectively.

Theorem 1: Let k be a real number and $\tilde{k}$ be defined as above. Also let ${Y_j}^{(u)}$ be as above, and $\tilde{Y}_j^{(u)} = \sum_{\forall i \in [m]} {A_i}^{(u)} \tilde{F}_{i,j}^{(u)}$, and $L = \sum_{\forall i \in [m]} |{A_i}^{(u)}|$.

Then for all $j = 1, 2, ..., g$, ${Y_j}^{(u)}$ and $\tilde{Y}_j^{(u)}$ are related as follows:

1) If ${Y_j}^{(u)} = k$,

$$(\tilde{k}^{(u)} - 0.5(1 + \sum_{\forall i \in [m]} |{A_i}^{(u)}|)) Q^{(u)} < \tilde{Y}_j^{(u)} < (\tilde{k}^{(u)} + 0.5(1 + \sum_{\forall i \in [m]} |{A_i}^{(u)}|)) Q^{(u)}$$

2) If ${Y_j}^{(u)} < k$,

$$\tilde{Y}_j^{(u)} < \tilde{k}^{(u)} Q^{(u)} + 0.5 Q^{(u)} (1 + \sum_{\forall i \in [m]} |{A_i}^{(u)}|)$$

3) If ${Y_j}^{(u)} > k$,

$$\tilde{Y}_j^{(u)} > \tilde{k}^{(u)} Q^{(u)} - 0.5 Q^{(u)} (1 + \sum_{\forall i \in [m]} |{A_i}^{(u)}|)$$

Theorem 2: For any quantizer scale $S \in [1, 31]$, the following condition is true.

$${Y_j}^{(u)} - \frac{95}{64} Q^{(u)} \sum_{\forall i \in [m]} |{A_i}^{(u)}| \leq \tilde{F}_{i,j}^{(u)} \leq {Y_j}^{(u)} + \frac{95}{64} Q^{(u)} \sum_{\forall i \in [m]} |{A_i}^{(u)}|$$

Then,

1) If ${Y_j}^{(u)} < k$, then $\tilde{Y}_j^{(u)} < k + \frac{95}{64} Q^{(u)} \sum_{\forall i \in [m]} |{A_i}^{(u)}|$.
2) If ${Y_j}^{(u)} = k$, ${Y_j}^{(u)} - \frac{95}{64} Q^{(u)} \sum_{\forall i \in [m]} |{A_i}^{(u)}| \leq \tilde{F}_{i,j}^{(u)} \leq {Y_j}^{(u)} + \frac{95}{64} Q^{(u)} \sum_{\forall i \in [m]} |{A_i}^{(u)}|$.
3) If ${Y_j}^{(u)} > k$, then $\tilde{Y}_j^{(u)} > k - \frac{95}{64} Q^{(u)} \sum_{\forall i \in [m]} |{A_i}^{(u)}|$.

1) Feature Code: A feature code is a binary string which represents ${Y_j}^{(u)}$; the length $N^{(u)}$ of a feature code is the accuracy (*precision*) of representation and determines the interval of the acceptable quality level for MPEG compression. Let $[{F_{MIN}}^{(u)}, {F_{MAX}}^{(u)})$ be the range of the DCT coefficient in position u, the sequence of bits $Z_{j,1}^{(u)}, Z_{j,2}^{(u)}, ..., Z_{j,N^{(u)}}^{(u)}$ represent the feature code for ${Y_j}^{(u)}$, and $U(N^{(u)}) = {F_{MIN}}^{(u)} + \sum_{i=1}^{N^{(u)}} ({F_{MAX}}^{(u)} - {F_{MIN}}^{(u)}) \cdot 2^{-i} \cdot Z_{j,i}^{(u)}$. Then the interval $[U(N^{(u)}), U(N^{(u)}) + ({F_{MAX}}^{(u)} - {F_{MIN}}^{(u)}) 2^{-N^{(u)}})$ satisfies the following condition [5].

$$U(N^{(u)}) \leq {Y_j}^{(u)} < U(N^{(u)}) + ({F_{MAX}}^{(u)} - {F_{MIN}}^{(u)}) 2^{-N^{(u)}}.$$

2) Finding the Tolerance Interval: The difference between ${Y_j}^{(u)}$ and $\tilde{Y}_j^{(u)}$ obtained from the decompressed image, is due to the *quantization error* and *calculation errors*.

The quantization error $\Delta = \tilde{Y}_j^{(u)} - {Y_j}^{(u)}$ is the weighted sum of m random variables, each corresponding to the quantization error of a single coefficient. That is, $\Delta = \sum_{i=1}^{m} {A_i}^{(u)} \delta_i$ where $\delta_i = \tilde{F}_{i,j}^{(u)} - {F_{i,j}}^{(u)}$. Our experiments showed that this variable has a symmetric zero-mean Gaussian-like distribution and its variance depends on the frequency and quality level of the compression. For lower quality levels and lower frequencies, the distribution of the quantization error has a large variance and is close to the uniform distribution.

The computation error is caused by inaccuracies introduced during computation, including the finite precision calculations used in the implementation of MPEG and other errors such as those resulting from integer representation of real valued

coefficients. Let $\varepsilon_{j,i}^{(u)} \in \mathbb{R}$ denote the error in the reconstructed value $\tilde{F}_{i,j}^{(u)}$. Then we have $\tilde{F}_{i,j}^{(u)} = rint(\frac{F_{i,j}{}^{(u)} + \varepsilon_{j,i}^{(u)}}{Q^{(u)}})Q^{(u)}$.

It can be shown [5] that the error in $\tilde{Y}_j^{(u)}$ is of the form $\sum_{\forall i \in [m]} A_i{}^{(u)} \varepsilon_{j,i}^{(u)}$. Let $\tau^{(u)}$ be a non-negative real number such that $-\tau^{(u)} \leq \sum_{\forall i \in [m]} A_i{}^{(u)} \varepsilon_{j,i}^{(u)} \leq \tau^{(u)}$ for $G_j{}^{(u)}, j = 1, 2, ..., g$ and assume $\varepsilon_{j,i}^{(u)}$ has a normal distribution with zero-mean. Then for large m,

$$\sum_{\forall i \in [m]} A_i{}^{(u)} \varepsilon_{j,i}^{(u)} \approx 0$$

and the errors will cancel out. However, using larger sums reduces security of the system.

B. Verification

Theorem 1 shows that the reconstructed linear sums will be close to the values calculated from the original frame. As long as $\tilde{Y}_j^{(u)}$ is within the interval $[U(N), U(N) + (F_{MAX}{}^{(u)} - F_{MIN}{}^{(u)})2^{-N})$, the verification is successful. The *verification tolerance* $E^{(u)} \geq 0$ for a given quality level ℓ is obtained by taking into account the quantization error and the computation error.

Verification proceeds as follows. If all feature codes pass the verification test, the image is considered authentic. To choose $E^{(u)}$ so that the quality level ℓ is acceptable, assuming the computation error $\tau^{(u)}$, we have $E^{(u)} \leq \frac{95}{64} Q_\ell^{(u)} \sum_{\forall i \in [m]} |A_i{}^{(u)}| + \tau^{(u)}$ where $Q_\ell^{(u)}$ is the maximum value of $Q^{(u)}$ corresponding to the lowest MPEG quality level that must be accepted by the system.

C. Authentication of P- and B-frames

Macroblocks in P- and B-frames can be encoded as *i)* motion vectors and error terms, or *ii)* intra-coded macroblocks and the method will depend on the bit-rate. The linear sums of DCT coefficients will be used to authenticate decoded P- and B-frames too. However, if the same method as I-frames is used for all frames, then the MAC size will be large. In the following, we consider methods of reducing the MAC size.

1) Reducing the MAC Size: Since P- and B-frames depends on I-frames, more protection needs to be provided for I-frames. To reduce the MAC size, we may *i)* reduce the number of frequencies which are protected; *ii)* reduce the number of bits used for each frequency; and *iii)* reduce the number of blocks used for generation of the MAC. We will mainly consider *iii)* as *i)* and *ii)* can always be used.

To reduce the number of blocks we may *i)* skip a frame completely, or *ii)* skip part of a frame. However these result in unprotected frames and regions that might be exploited by the attacker.

A video stream consists of a sequence of pictures. To produce a visible change in the video, the attacker must modify a number of consecutive frames. To protect a video stream against such attacks, it is sufficient to verify frames at a high enough rate such that the visible changes become detectable. Assuming changes to frames will be applied on consecutive frames for at least 60 millisecond, we may leave out some of the frames from the signature altogether. For example, in PAL and NTSC systems, there are about 30 frames/second, and so each frame will be 30 milliseconds. This means that the MAC needs to be calculated for every other frame.

An alternative to skipping a whole frame is to choose a subset of blocks in each frame such that the union of the subsets in consecutive frames cover the whole frame. For example we may choose subsets as checkerboard pattern such that alternate blocks are chosen in two consecutive frames. This method will provide good protection assuming no major change occur between two consecutive frames.

D. SARI Authentication System

Lin and Chang [6] proposed an image authentication system known as SARI, that was later extended to video authentication [7], [8]. SARI can be considered as a special case of the proposed system where $m = 2$ and $A_1{}^{(u)}, A_2{}^{(u)}$ are $1, -1$, respectively, and so $\sum_{\forall i \in [m]} |A_i{}^{(u)}| = 2$.

IV. Keyed Hash Function

Feature codes can be seen as hash values of a frame. However, if all design parameters of the system are public, it will be easy to construct two frames with the same feature code. To provide collision resistance, some key information must be introduced. The result will be *a keyed hash function* or a *message authentication code (MAC)*. The key information will be kept secret and shared by the authentication and the verification systems. Parameters of the system may be used as the key are; *i)* m, the number of blocks in a group; *ii)* The composition of groups, $G_j{}^{(u)}$, $j = 1, 2, 3, ..., g$; *iii)* S, the set of protected frequencies u; *iv)* Coefficients of the linear combination $A_i{}^{(u)}$, $i = 1, 2, 3, ..., m$; *v)* The precision (number of bits) $N^{(u)}$ of the feature code for frequency u; and *vi)* The error tolerance $E^{(u)}$.

The composition of groups can be specified by an incidence matrix, whose rows correspond to groups and columns correspond to blocks. The matrix entries are 0 and 1 with 1 in (i, j) position showing that group i includes block j. We assume groups have the same size. A larger size for a group gives more flexibility to the attacker to modify a target block and spread the compensating change over the rest of the blocks in the group. On the other hand larger groups mean that less groups are needed to cover the whole image and so shorter MAC will be produced. The number of blocks in a group must be chosen by taking these conflicting requirements into account. Typical values are 8, 16, and 32.

If groups are disjoint, the change in a block will affect one feature code. To spread the change in a block to other blocks, groups must be *linked*. Two groups G_i and G_j are *linked* if there is a sequence of groups $G_i, G_{i+1} \cdots G_j$ such that $G_t \bigcap G_{t+1} \neq \emptyset, t = i, i+1, \cdots, j$ where $i < j$. We require that groups be linked.

1) Linear Combination Coefficients: Let $A_{MIN}{}^{(u)}$ and $A_{MAX}{}^{(u)}$ denote the maximum and the minimum value of the linear combination coefficients and assume $A_i{}^{(u)}$ is chosen

randomly from the interval $[A_{MIN}{}^{(u)}, A_{MAX}{}^{(u)}]$. A change α in a DCT coefficient is multiplied by $A_i{}^{(u)}$ and so will be magnified for high multipliers. If all regions of the image have the same significance, then $A_i{}^{(u)}$ must be chosen equal or close to each other.

2) A Proposal for MAC: We choose the secret key information to be the mapping between the the incidence matrix and the blocks in a frame. This means that the size of the key space is $\wp!$. The system parameters such as precision of coefficients, the number of groups, and the protected frequencies will be determined by other considerations such as the length of the MAC, and the quality level of the MPEG compression that must be tolerated.

V. Evaluation of the MAC

3) Security: We show an estimate of the cost of constructing a fraudulent frame that passes the verification test. The cost of constructing a fraudulent video clip by using the best known attack [9] is lower bounded by this. The analysis of the system is using the following scenario.

An attacker has access to a video decoder which contains the key in a tamperproof device. He wants to modify a single frame in an authentic MPEG stream. The attacker can input various frame and MAC pairs and receive the response of the decoder.

He will succeed if he can construct a frame and MAC pair that is acceptable by the decoder. He does not know the blocks in each group. To modify a chosen block the attacker has to find other blocks in the same group and modify them such that the change to the chosen block is compensated. If groups are linked, a change propagates through all groups.

It can be shown [5] that the cost of finding all groups is at most $\sum_{j=1}^{g} \binom{\wp-(m-1)(j-1)}{m}$ verification operations and the average is $(\sum_{i=1}^{g}\sum_{j=1}^{i} \binom{\wp-(m-1)(j-1)}{m})/g$. For example, if $\wp = 4096$ and $m = 8$ then $gm = 2\wp$ because each block belongs to two groups and so $g = 1024$ and the cost of finding all blocks of $G_j, \forall j$ is $C \approx 10^{26}$ verification operations.

A similar attack can be used against the SARI system [9]. The costs of finding a pair of blocks, and all pairs, are $\wp - 1$ and $\wp^2/4$, respectively. For example, if $\wp = 4096$, finding all pairs costs about 4×10^6, which is much smaller than $C \approx 10^{26}$ and so SARI is considered to be insecure.

4) Efficiency of the MAC: The length of the MAC for one frame is given by $L = g\sum_{u\in S} N^{(u)}$ where S is the set of protected frequencies. The length of the MAC is proportional to the number of groups. This suggests to have larger number of blocks per group. The number of bits allocated to the feature codes corresponding to $Y_j{}^{(u)}$ is determined by the compression level that must be tolerated. If only high quality images must be acceptable, then more bits must be allocated to the feature codes.

5) Implementation and Experiments: We implemented the systems and performed a number of experiments to verify our results. These experiments show that local and global modifications can be effectively detected for MAC sizes of 8K per frame for a 352×240 color MPEG stream. However small modification may remain undetected if smaller MAC sizes are used. To examine tamper detection property of the system, we used a stream "table tennis" (Fig.1). The number "31.47.48" on the poster was modified to "41.47.48." in an I-frame (PSNR Red: 37.0 dB, Green: 37.0 dB, Blue: 38.3 dB). The modification was detected by the system.

Fig. 1. Tamper detection experiment: The original (left) and the modified frame (right).

VI. Conclusion

We proposed an MPEG tolerant video authentication system with adjustable security, that can be efficiently incorporated into the MPEG compression algorithm. The MAC consists of a number of feature codes that can be carefully chosen to protect regions of interest in the image and to various levels of accuracy. MAC generation and verification are very efficient and assuming the DCT coefficient are available, only requires a small number of multiplication and addition.

We analyzed security of the MAC in a proposed attack scenario and estimated the computational cost of attack under the best known strategy. More formal modeling and analysis of security remains a challenging open problem.

References

[1] Win Wu and Bede Liu, "Watermarking for image authentication," in *Proc. IEEE Int. Conf. on Image Processing*, 1998, pp. 437–441.

[2] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete Cosine Transform," *IEEE Trans. on Computers*, vol. C-23, pp. 90–93, 1974.

[3] ISO/IEC, *MPEG Standard*, http://www.mpeg.org, 1998.

[4] David Salomon, *Data compression : the complete reference 2nd edition*, Springer-Verlag NewYork, Inc., 2000.

[5] Takeyuki Uehara, Reihaneh Safavi-Naini, and Philip Ogunbona, "A secure and flexible authentication system for digital images," *Journal of ACM Multimedia Systems*, vol. 9, no. 5, pp. 441–456, March 2004.

[6] Ching-Yung Lin and Shih-Fu Chang, "Robust image authentication method surviving JPEG lossy compression," *Proc. of SPIE Storage and Retrieval for Image and Video Databases*, vol. 3312, pp. 296–307, Jan 1998.

[7] Ching-Yung Lin and Shih-Fu Chang, "Issues and solutions for authenticating mpeg video," in *Proc. of SPIE Security and Watermarking of Multimedia Contents*, Jan 1999, vol. 3657, pp. 54–65.

[8] Ching-Yung Lin and Shih-Fu Chang, "SARI : Self-authentication-and-recovery image watermarking system," in *Proc. of the ninth ACM international conference on Multimedia*, 2001, pp. 628–629.

[9] Takeyuki Uehara and Reihaneh Safavi-Naini, "On (in)security of "a robust image authentication method"," in *Proc. of the 2002 IEEE Pacific-Rim Conference on Multimedia*. Dec 2002, vol. LNCS2532, pp. 1025–1032, Springer-Verlag.