A CONTENT-AWARE STREAM AUTHENTICATION SCHEME OPTIMIZED FOR DISTORTION AND OVERHEAD

Zhishou Zhang[1,2], Qibin Sun[1], Wai-Choong Wong[1,2], John Apostolopoulos[3] and Susie Wee[3]

[1]Institute for Infocomm Research, Singapore
[2]Department of ECE, National University of Singapore, Singapore
[3]Hewlett-Packard Laboratories, Palo Alto, CA USA

ABSTRACT

This paper proposes a content-aware authentication scheme optimized to account for distortion and overhead for media streaming. When authenticated media is streamed over a lossy network, a received packet is consumed only when it is both decodable and authenticated. In most media formats, some packets are more important than others. This naturally motivates allocating more redundant authentication information for the more important packets in order to maximize their probability of authentication and thereby minimize distortion at the receiver. Toward this goal, with awareness of the media content, we formulate an optimization framework to compute an authentication graph to maximize the expected media quality at the receiver, given specific authentication overhead and knowledge of network loss rates. Experimental results with JPEG-2000 coded images demonstrate that the proposed method achieves our design goal in that the R-D curve of the authenticated image is very close to the R-D curve when no authentication is required.

1. INTRODUCTION

Media streaming is gaining in importance, however security issues have to be addressed to facilitate secure exchange of media, e.g. secure adaptation of secure digital imagery [1]. This paper deals with integrity and source authentication issues for media streaming over a lossy network.

Common approaches are to amortize one signature among a group of packets, which are connected via a directed acyclic graph. Each node corresponds to a packet and an edge corresponds to a crypto hash link. For example, an edge from packet *A* to *B* is implemented by appending *A*'s crypto hash to *B*. The graph typically has only one packet carrying the signature, which is referred to as the signature packet. Each packet has at least one path to the signature packet. At the receiver, the lost packets are removed from the graph, and a packet is verifiable, and therefore decodable, only if it still has a path to the signature packet. A packet with more redundant paths will have higher verification probability, but at the cost of more overhead from the associated hashes. Therefore, minimizing hash overhead and maximizing verification probability are competing goals. Existing stream authentication methods [2]-[6] are designed to achieve a balance between these two competing goals.

However, for media streams, we argue that the authentication schemes should be measured by media quality instead of verification probability, because the quality of the final authenticated media is what matters. Therefore, media quality and overhead should be used as metrics for evaluating authentication methods. Furthermore, existing authentication schemes either implicitly or explicitly assume that all packets are equally important, which is typically not true for media. Therefore authentication schemes for media streaming which account for distortion and overhead are needed.

In this paper, we propose a content-aware authentication scheme optimized for distortion and overhead for media streaming. We define that the received media is authentic if it is exclusively decoded from received packets that have passed verification. As the proposed scheme differentiates the packets by their importance, it is able to optimally allocate authentication overhead to individual packets in order to minimize the expected distortion at the receiver. For example, more important packets have their hashes replicated and appended in greater number to other packets, which increases their probability of being verifiable.

This paper continues by describing the proposed content-aware media stream authentication method in Section 2. Section 3 analyzes system and security issues. The proposed scheme is validated in Section 4 by experimental results.

2. PROPOSED OPTIMIZED STREAM AUTHENTICATION SCHEME

We assume a general layered media format shown in Fig. 1, of *M* independent coding units where within each unit a packet depends on the corresponding lower layer packets for decoding (linear dependency). For instance, P_m^l is decodable if all the packets from P_m^0 to P_m^{l-1} are received. Associated with P_m^l is the distortion reduction ΔD_m^l, the amount by which the overall distortion will reduce if it is decoded. Here we assume distortion to be additive, which is a simplified distortion model for packet loss.

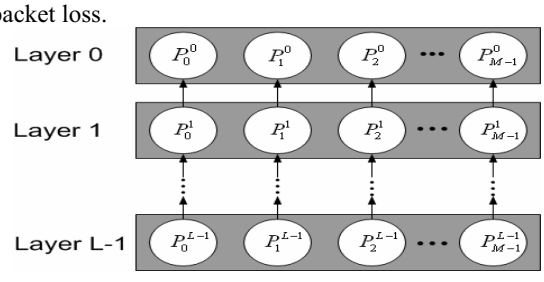

Fig. 1 – Media format with *M* units and *L* layers

2.1 Distortion-overhead optimized authentication graph

The packets are connected by an authentication graph denoted by <*V,G*>, where *V* is the set of packets and *G* is the set of edges. In an edge from packet *A* to *B*, *A* is called a *source packet* and *B* is called a *target packet*. Let π_m^l be the *target set* consisting of target packets of the edges coming out of P_m^l, and $|\pi_m^l|$ is the *redundancy degree* of P_m^l. Then we can use the vector $\pi=[\pi_0^0, \pi_1^0, ..., \pi_m^l, ..]$ to uniquely define the authentication graph, given the set of packets *V*. Denote the total authentication overhead (i.e., digital signatures and crypto hashes) as $O(\pi)$ and expected overall distortion as $D(\pi)$, our goal is to find the optimal vector π that minimizes the Lagrangian cost function shown in (1), given $\lambda>0$.

$$\pi^* = \arg\min_{\pi} \left(D(\pi) + \lambda O(\pi) \right) \tag{1}$$

The authentication overhead $O(\pi)$ includes hashes appended to packets and the signature, as calculated in (2), where SIZ_{sig} and SIZ_{hash} are the sizes of the signature and hash respectively.

$$O(\pi) = SIZ_{sig} + \sum_{P_m^l} \left| \pi_m^l \right| SIZ_{hash} \tag{2}$$

The expected distortion $D(\pi)$ can be computed in (3), where D_0 is the distortion when no packet is decoded, ρ_m^l denotes the probability that P_m^l is decodable, and 1-$\varepsilon(\pi_m^l)$ denotes the conditional verification probability with π_m^l given that P_m^l is decodable.

$$D(\pi) = D_0 - \sum_{m=0}^{M-1} \sum_{l=0}^{L-1} \Delta D_m^l \rho_m^l \left(1 - \varepsilon(\pi_m^l) \right) \tag{3}$$

Determining the optimal vector π is accomplished in two stages: the first stage is to determine π_m^l when $l>0$, while the second stage is to determine π_m^l when $l=0$.

Stage 1: For packet P_m^l when $l>0$, it is sufficient to have one edge from P_m^l to its immediate ancestor, i.e. $\pi_m^l = \{P_m^{l-1}\}$, which has the minimal overhead (1 hash) and maximal conditional verification probability ($\varepsilon(\pi_m^l)=0$).

Stage 2: After stage 1, ρ_m^l can be expressed in (4), where ϕ_m^i is probability that P_m^i is received. In other words, P_m^l is decodable if and only if all its ancestors (i.e. $P_m^0, P_m^1, \ldots, P_m^l$) are received and P_m^0 is verifiable.

$$\rho_m^l = \left(1 - \varepsilon(\pi_m^0)\right) \prod_{i=0}^{l} \phi_m^i \tag{4}$$

Substituting (4), (3) and (2) into (1), we can get the Lagrangian cost function with variables $\varepsilon(\pi_m^0)$ and $|\pi_m^0|$, where $0 \le m < M$. This problem can be solved in an iterative way, i.e. the Lagrangian cost function is minimized by searching for the optimal *target set* for one packet at a certain iteration, keeping the other packets' *target set* unchanged, until the cost function converges. For instance, at certain iteration, for packet P_m^0, the optimal *target set* π_m^0 can be decided by (5). To ensure the authentication graph is acyclic, we mandate that for edges whose source packet is P_m^0, the target packet must be P_n^0 where $n<m$, as any cycle in an authentication graph will make it impossible to compute the hash appended to the signature packet.

$$\pi_m^0 = \arg\min_{\pi'} \left(\varepsilon(\pi') + (\lambda |\pi'| / \mu_m^0) SIZ_{hash} \right) \tag{5}$$

where the utility value μ_m^0 is expressed in (6).

$$\mu_m^0 = \sum_l \left(\Delta D_m^l \prod_{i=0}^{l} \phi_m^i \right) \tag{6}$$

The utility value μ_m^0 can be considered as the amount by which the distortion will increase if P_m^0 is not verifiable. Therefore, the larger the utility value μ_m^0 is, the more attractive it will be to increase the verification probability of P_m^0.

2.2 Simplified authentication graph

Section 2.1 computes an optimized authentication graph, which is computational-intensive since many iterations are required before convergence. Here we present an approach to compute a simplified authentication graph with much lower complexity.

For P_m^l when $l>0$, there is only one edge from this packet to its immediate ancestor, i.e. $\pi_m^l = \{P_m^{l-1}\}$. For P_m^0, we compute the utility value μ_m^0 using (6). All layer-0 packets are sorted in descending order of utility in a list, which is then divided into *S* segments of equal size, namely Seg_0, Seg_1,…,Seg_{S-1}. Each packet in Seg_i has γ_i outgoing edges whose target packets are randomly selected from the preceding packets in the sorted list. The redundancy degree for different segments is in non-increasing order, i.e. $\gamma_0 \ge \gamma_1 \ge \ldots \ge \gamma_{S-1}$. The signature packet contains hashes of the first *Z* packets in the sorted list.

3. PERFORMANCE ANALYSIS

3.1 Comparison with existing schemes

Since the proposed method has a directed acyclic graph structure, we choose five existing graph-based authentication methods for comparison, including simple hash chain [2], Tree-authentication [3], EMSS [5], Augmented Chain [4] and Butterfly [6]. The benchmark criteria are summarized as follows:

1) *Computation overhead*: number of crypto hash operations (e.g., SHA-1 [7]) and signature signing / verifying operations required at the sender and receiver.
2) *Communication overhead*: number of extra bytes carried by each packet for authentication.
3) *Verification percentage* number of verifiable packets divided by the number of received packets.
4) *Sender delay*: delay in packets at the sender before the first packet can be transmitted.
5) *Receiver delay*: delay in packets at the receiver before the first packet can be verified and decoded.

In terms of computation overhead, Tree-authentication has to perform about *N* more hash operations than the other schemes. In terms of sender delay and receiver delay,

Augmented Chain has the worst performance, while the other schemes are the same.

Simple Hash Chain has the smallest communication overhead (one hash per packet), but its verification probability is the lowest. On the other hand, Tree-authentication has the highest communication overhead, but also the highest verification probability. These two schemes are the extreme cases (one favors communication overhead while the other favors verification probability), while all other methods try to achieve a balance in-between. In particular, both EMSS and our proposed content-aware scheme can be configured with different overhead levels, resulting in different verification probabilities. The Augmented Chain and Butterfly method have fixed communication overhead. In this regard, EMSS and our scheme are more flexible and generically applicable since their overhead levels are tunable.

The above comparisons are for streaming general data packets. However, for media streams we should use different benchmark criteria. The most important measure should be the quality of the authenticated media rather than the verification probability of the delivered packets. In this regard, our content-aware scheme is more efficient than existing schemes due to two reasons:

1) It eliminates all unnecessary hash links that only help to increase the verification probability, but not the PSNR of the authenticated image. For example, if packet *A* depends on *B*, it is sufficient to have an edge from *A* to *B*, and an edge from *A* to any other packet does not help to improve the PSNR of the authenticated image.
2) The edges are used in a more efficient manner. The more important packets own more outgoing edges, which help to increase the PSNR, while the less important packets (which account for a large proportion of the total packets) have less outgoing edges resulting in smaller communication overhead.

3.2 Security Analysis

Similar to existing graph-based authentication schemes, the content-aware authentication relies on the hash chain and digital signature. Therefore, the security strength of this scheme is the same as the underlying cryptographic algorithms. For example, SHA-1 [7] can be used for one-way hashing and DSA [7] can be used for signature generation and verification. For more details on security strength analysis, interested readers please refer to Merkle's tree authentication [8].

3.3 Discussion on utility values

As every network has a maximum transmission unit (MTU) size, a packet has to be segmented into smaller datagrams for transmission. Therefore, the probability of receiving the packet P_m^l, ϕ_m^l, can be expressed in (7), assuming network loss follows i.i.d distortion, p_{lost} is average loss probability of the network datagram and R_m^l is the packet size in bytes.

$$\phi_m^l = (1 - p_{lost})^{\lceil R_m^l / MTU \rceil} \quad (7)$$

By substituting (7) into (6), we can see that the utility value μ_m^0 is determined by the associated distortion ΔD_m^l, packet size R_m^l of its descendent packets, the network loss probability p_{lost}, and *MTU*. If the packet size R_m^l is greater, it needs more datagrams for transmission, decreasing its probability of being received. So, given a fixed value for p_{lost} and *MTU*, the packet P_m^0 will have a greater utility value when its corresponding higher-layer packets (including P_m^0 itself) have larger ΔD_m^l and smaller R_m^l.

The *MTU* value depends on the physical network links, e.g. Ethernet has a *MTU* of about 1500 bytes and the ATM network has a *MTU* of 53 bytes. The sender could presume a reasonable value for the loss probability p_{lost}, or could estimate it based on past communications. However, for the simplified authentication graph, it is not important to have an accurate value for p_{lost}, because it does not change the relative order of the utility values in the sorted list.

4. EXPERIMENTAL RESULTS

This section experimentally compares our content-aware scheme against EMSS, Augmented Chain and Butterfly authentication schemes using JPEG-2000 images to further demonstrate the validity of our proposed scheme.

We implemented five schemes based on JPEG-2000 [9]. The first scheme, *WITHOUT_AUTH*, does not apply any authentication. It provides a reference for the achievable distortion performance if verification is not required. The second scheme, *EMSS_AUTH*, implements EMSS. The third scheme, *CONTENT_AUTH*, implements our proposed content-aware authentication using the simplified authentication graph. Through simulation, we find that the content-aware scheme yields good performance using segmentation with S = 3. Further increasing S does not produce substantial performance improvement, because its performance is already quite close to the upper bound when S is set to 3. The fourth scheme, *AC_AUTH*, implements the Augmented Chain, and the fifth scheme, *BUTTERFLY_AUTH*, implements the butterfly authentication. For all schemes, the packets are sent in the order they appear in JPEG-2000 codestreams, while the signature packet is sent multiple times to minimize its loss probability.

The network is modeled by an i.i.d distribution, where the average loss probability ranges from 0 to 0.15. In addition, *MTU* is set to 1500 bytes, as used by Ethernet. Our experiment uses the 8 JPEG-2000 testing images (each 2560x2048 pixels) and each image has 260 packets per layer.

The first experiment demonstrates the effectiveness of the authentication redundancy adapted to the distortion. The JPEG-2000 images are encoded with only 1 layer, so *CONTENT_AUTH* can take advantage of the distortion information but not the layer structure. For *CONTENT_AUTH*, the parameters are set as follows: S=3, $\gamma_0 = 3$, $\gamma_1 = 2$, $\gamma_2 = 1$, and Z=6, so the redundancy degree is 2 on average. Similarly, the other schemes use the similar level of redundancy, i.e. redundancy degree is 2. Fig. 2 shows the PSNR of the schemes. *CONTENT_AUTH* consistently outperforms the other schemes at all network loss rates. In fact, the PSNR curve of *CONTENT_AUTH* is very close to that of *WITHOUT_AUTH*,

which achieves our original design goal, because the authentication overhead is added in an optimized manner.

The second experiment evaluates the proposed scheme when both distortion and layer structure are utilized. Accordingly, the JPEG-2000 images are encoded with 6 layers. Thus, *CONTENT_AUTH* is able to take advantage of both the layer structure and the distortion information, but other schemes are agnostic to both. The parameters for this experiment are the same as that for the previous experiment. Fig. 3 shows the PSNR curves of the three schemes, which are similar to those in Fig. 2.

The third experiment compares the performance at various overhead levels. We set the loss probability to 0.05, which is usual for many scenarios. The JPEG-2000 images are encoded with 1 layer, because we want *CONTENT_AUTH* to utilize the distortion information but not the layer structure. We measure the PSNR at various overhead levels, ranging from 1 to 6 hashes per packet. Fig. 4 shows that at loss rate 0.05 the proposed scheme outperforms the other schemes when the redundancy degree is less then 3. When the loss rate is higher, the gap between the proposed scheme and the other schemes will be further increased.

The fourth experiment measures the minimum overhead required to achieve a PSNR that is 99% of *WITHOUT_AUTH*. The JPEG-2000 images have one layer, for the same reason as the previous experiment. As shown in Fig. 5, *CONTENT_AUTH* requires an overhead of 2 hashes/packet when the loss rate <= 0.1, and requires 3 hashes/ packet for 0.1 < PLR <= 0.15. However, *EMSS_AUTH* requires more overhead in order to maintain the same PSNR level.

5. CONCLUSIONS

This paper proposed a content-aware authentication scheme, which achieves distortion-overhead optimization by utilizing the hints from media content and layer structure. In view that the optimization process has high computational complexity, we also proposed a lower-complexity stream authentication graph. Experimental results have demonstrated that the PSNR curve of the content-aware authentication method is very close to the upper bound achievable when authentication is not required, and it substantially outperforms existing schemes of content-unaware authentication.

6. REFERENCES

[1] S. Wee and J. Apostolopoulos, "Secure transcoding with JPSEC confidentiality and authentication," in Proc. IEEE International Conference on Image Processing, October 2004

[2] R. Gennaro and P. Rohatgi. "How to sign digital streams," in Advances in Cryptology - CRYPTO '97, pp. 180–197.

[3] C. K. Wong and S. Lam, "Digital Signatures for Flows and Multicasts", The University of Texas at Austin, Department of Computer Sciences, Technical Report TR-98-15. July 1998

[4] P. Golle and N. Modadugu. "Authenticating streamed data in the presence of random packet loss," ISOC Network and Distributed System Security Symposium, 2001, pp 13--22.

[5] A. Perrig, R. Canetti, J. Tygar and D. Song. "Efficient authentication and signing of multicast streams over lossy channels," in Proc. of IEEE Symposium on Security and Privacy, 2000, pp. 56-73.

[6] Z. Zhang, Q. Sun and W-C Wong, "A Proposal of Butterfly-based Stream Authentication Scheme over Lossy Networks," in Proc. of International Conf. on Multimedia & Expo, 2005

[7] B. Schneier, Applied Cryptography, Wiley, 1996.

[8] R. C. Merkle, "A certified digital signature," in Advances in Cryptology – CRYPTO'89, 1989, pp. 218-238

[9] Information technology – JPEG2000 image coding system, ISO/IEC International Standard 15444-1, ITU Recommendation T. 800, 2000

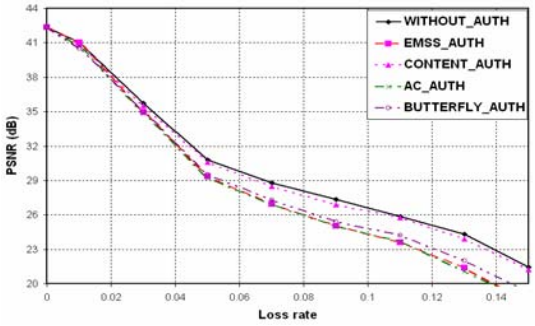

Fig. 2 – PSNR versus loss rate (2 hashes/packet, with 1 layer)

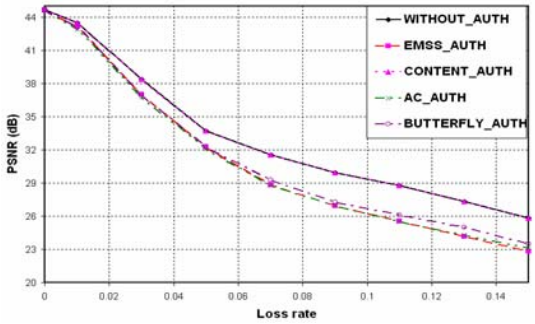

Fig. 3 – PSNR versus loss rate (2 hashes/packet, with 6 layers)

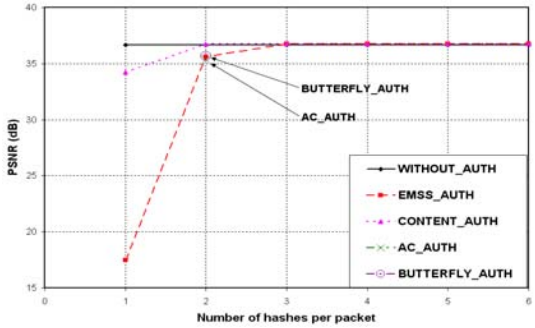

Fig. 4 – PSNR versus overhead (loss = 0.3, with 1 layers)

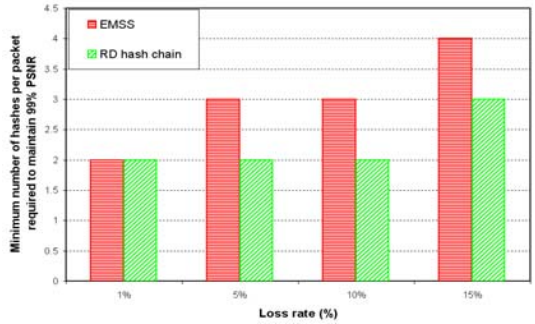

Fig. 5 – Minimum overhead required to achieve 99% PSNR at various rates (with 1 layer)