Secure Video Multicast over Wireless Ad-hoc Networks using Network Coding

Du Yang, Valdemar Monteiro and Jonathan Rodriguez
Instituto de Telecomunicações
Universidade Aveiro, Campus Universitário,
Aveiro, 3810-193, Portugal
vmonteiro@av.it.pt

Tasos Dagiuklas and Charalambos Mysirlidis
Hellenic Open University, Computer Science
Patras, 263 65, Greece
ntan@ece.upatras.gr

***Abstract*—Video streaming over wireless networks has been continuously increasing, which results in significant energy consumption and growing security concerns, especially for safety-related services such as video surveillance. Advanced transmission and security mechanisms are required to improve video quality, protect video transmissions, and in the same time achieve energy efficiency. One of the promising solutions is to exploit network coding. In this research paper, we provide network coding aided transmission as well as security mechanisms, and compare them to non-network coding schemes. It is proved that using network coding is capable of significantly improving the achievable video quality, and reducing the computation complexity as well as signaling overhead for data encryption.**

Keywords-* *secure video streaming; network coding; SVC.

I. INTRODUCTION

Media Applications such as broadcasting IPTV video and video on demand (VoD) services have become extremely popular to both service providers that perceive them as a mean to expand their revenues and market share, and subscribers that can have access to all-IP based services and traditional TV on any device, through the Internet. The H264/SVC (Scalable Video Coding) as specified in Annex G of H.264/AVC allows the construction of bit-streams that contain sub-bit streams that conform to H.264/AVC [1].

Network Coding is a promising technique that could be used for network content distribution. The concept behind network coding relies on XOR/linear combination among the packets in order to optimize throughput [2]. However, one of the constraints is because it necessitates considering coding distortion conveyed in a video packet in order to construct the network information flows in an efficient way [3].

The wireless media but other hand, is vulnerable for a wide range of attacks because of its broadcasting nature, which motivates a continuous research interest in this area [4][5] with the objective of providing sufficient security protection in an resource efficient manner.

This paper examines the use of secure network coding for video applications over wireless environments. It is proved that using network coding is capable of achieving the max-min-cut network capacity [6], and providing weak information-theoretical security [7].

The rest of the paper is organized as follows. The target scenario is introduced in section II, followed by a brief introduction of scalable video coding and network coding in section III. Video transmission schemes on multi-hop wireless network with and without network coding are detailed in section IV. In section V, two network coding assisted security mechanisms are described. The simulation results are demonstrated in section VI, followed by conclusion and future work in section VII.

II. TARGET SCENARIO

We are interested in considering an ad-hoc network, where nearby nodes are organized into clusters. One of the trustworthy nodes such as a desktop/router is selected as the cluster head. This cluster head is responsible of connecting to other cluster heads or severs. The example two-hop network as shown in Figure 1 consists of one source node (Node-S), three relay nodes (Node-R1, Node-R2 and Node-R3), and two destination nodes (Node-A and Node-B). The wireless links are denoted in dashed lines, which results in three disjoint paths from source node to each destination node via three different relay nodes. All the nodes are connected to other clusters or servers via the cluster-head R2.

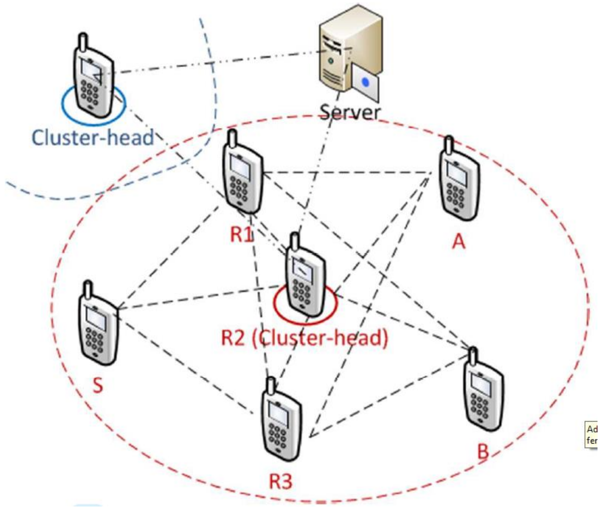

Figure 1. Target scenario example: a two-hop wireless ad-hoc network consisting of five nodes.

Besides serving as a portal, the cluster-head is also evolved in key management. At the beginning of clustering, the cluster head broadcasts its presence. Each member node responses with a joining request, and negotiate a shared key using Diffie-Hellman key exchange algorithm [8].

Our target is to multicast a video sequence within this network in an efficient and secure manner. Without loss of generality, we focus on the communication within a single cluster. As shown in Figure 1, a H.264/SVC encoded video sequence is generated at Node-S, and transmitted to Node-A and Node-B. We will demonstrate in the rest of the paper that network coding could be applied in both transmission and security mechanisms so as to improve the efficiency.

III. NETWORK CODING AIDED TRANSMISSION MECHANISMS

In this section, we start introducing some preliminary concepts, followed with introducing the common procedure of video transmission. Then we detail in single hop scenario, how the video packets are transmitted without and with network coding. After that, we discuss the additional benefits of using network coding in a multi-hop scenario.

A. *Scalable Video Coding and Network Coding concepts*

There are different ways of introducing scalability in H.264 SVC. The bit stream supports the following scalability modes: Coarse-Grain Scalability (CGS: the transform coefficients are encoded in a non-scalable way), Medium Grain Scalability (MGS the transform coefficients can be split in several fragments) and Fine-Grain Scalability (FGS: the transform coefficients are arranged as an embedded bit stream). Without loss of generality, MGS has been used in this research work.

Network coding is a revolutionary idea proposed in [6], which considered information bits as flow instead of commodity. It allows intermediate nodes in the network to store-recode-forward received information, instead of simply store-forward. Network coding algorithms can be classified according to various criteria [9]. Based on the present/absent of network topology knowledge, there are state-aware/stateless network coding algorithms.

The delay-sensitive nature of video streaming requires fast en-/decoding operation. Considering these two factors, stateless random linear network coding (RLNC) and the simply XOR operation have been widely recognized as good candidates [3,5,9] for video multicast over wireless network, which are also deployed in this paper.

B. *Common video transmission procedure*

A general video transmission involves three entities: source node, network, and sink node. Video is sampled at the source node frame by frame with a constant sampling interval Tf(s). Each frame is compressed and encoded into a sequence of packets. As demonstrated in Figure 2, each frame is encoded into two layers of packets, a base-layer ($l = 1$) and an enhancement layer ($l = 2$) (Without loss of generality, we assume 2 layers generated using MGS). It is crucial for all destinations to receive base-layer packets. The enhancement layer packets are optional and not decodable without receiving the base-layer.

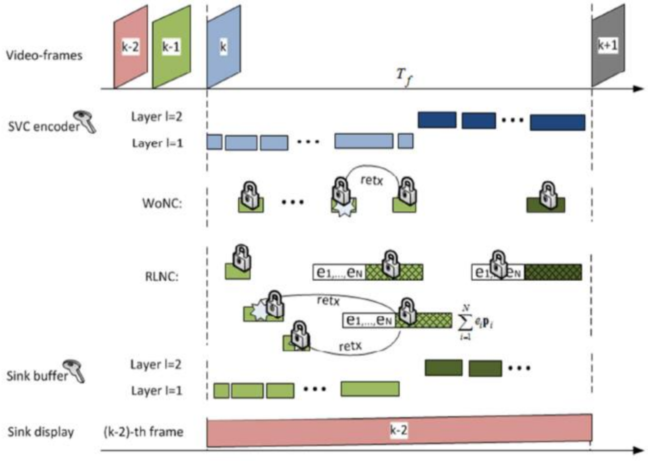

Figure 2. Illustration of video transmission and data encryption 1) without network coding; and 2)with network coding in a single-hop scenario.

It is essential for a sink node to buffer a few packets or frames before replaying the video. At a certain time slot, source node is in the process of compressing and encoding the k -th frame. Concurrently, packets belonging to the (k −1) -th video frame are transmitted and retransmitted over the network, and buffered at sink nodes. In the meantime, the (k − 2) -th video frame is displayed. As a result, the total transmission and retransmission time for each frame is equal to Tf.

The H.264 SVC bit stream is organized in autonomous entities called NAL Units. Each NAL Unit contains a header and payload. NAL units are encapsulated in RTP packets. IETF has specified different modes of encapsulating NAL Units in RTP packets namely: Single NAL Unit (SNU), Aggregating NAL Unit from one frame to RTP (STAP), Aggregating NAL Units from different frames to RTP (MTAP), Fragmenting NAL Units to Multiple RTP Packets. There are 3 transmission modes for SVC namely: Single Session Transmission (SST), Multi-Session Transmission (MST) and Media-Aware Network Element (MANE). Without loss of generality, we use SNU for each layer and MST SVC Transmission Mode (Each layer sends independently its packets). We are interested in using network coding to reduce the packet loss ratio of RTP video packets, in order to improve the video quality.

C. *Single-hop scenario*

The transmission in a single-hop scenario is illustrated in Figure 2 without and with network coding.

Without network coding scheme (WoNC): In the absent of network coding, the RTP video packets capsulated into IP packets by appending the IP header, and then pass to media access control (MAC) layer. Whenever a mobile node seizes an opportunity to transmit, adaptive modulation and coding is applied on the MAC layer packets based on the channel condition. An acknowledgement is fed back to the transmitter when a packet is received. Otherwise, this MAC-layer packet is marked as lost, and will be retransmitted later. Note that although end-to-end retransmission is forbidden in RTP/UDP/IP protocols, single-hop retransmission at MAC layer is allowed.

With RLNC: When we have equal length data packets, RLNC scheme logically re-organized the original packets { p1, p2,... } at source nodes into generations, each of which is a set of packets with adjacent packet-ids. We use a pair of parameters (g_{id} , g_{size}) to denote a generation to which packets with packet-ids equal to or larger than (g_{id} -1)*g_{size} , and smaller than $g_{id}\, g_{size}$. A coded packet is generated by linear combination, where e_k is an element in a certain finite field F. In the header of a coded packet, the encoding vector e = (e_1, K, e_{size}) as well as g_{id} is stored for later decoding at the receivers. A destination must receive equal or more than g_{size} number of linear independent coded packets in order to decode the original packets via Gaussian elimination.

However, SVC video packets have several features very distinct from non-real-time data packets. First of all, the video packets do not have equal length, but varies from several bytes to 1000 bytes according to frame type and scene complexity. Simple padding will introduce a significant amount of overhead. Second, video service is sensitive to delay, which could not tolerant a big generation size. Third, some video packets are more important than others. For example, base-layer video packets are more important than enhancement-layer ones. Mixing those packets with different importance may introduce video quality degradation.

Considering the aforementioned video features, we improve the network coding algorithm proposed in [3][10][11], and propose the following alterations:

- Base-layer video packets are first transmitted without using network coding.
- During base-layer packet retransmission, RLCN is applied over several packets. Supposed that four packets {p_1,p_2,p_3,p_4} are transmitted from Node-S. Only p_2 is lost at Node-R1, and p_3 is lost at Node-R2. Both nodes can recover their lost packet if Node-S retransmits one coded packet $\sum_{i=1}^{4} e_i \mathrm{p}_i$
- Enhancement layer video packets are transmitted and retransmitted using RLCN, since they usually have a longer packet length and smaller variation.
- Enhancement layer video packets can be encoded with base layer video packets so as to provide higher protection to base-layer packets. Packet length variation is coped with the generation size allowed to change.

As we demonstrate in the simulation section, the proposed RLNC scheme reduces the Packet Loss Ratio (PLR) compared to WoNC scheme, and improves the video quality.

D. Multi-hop scenario

Using network coding provides an additional benefit in a multi-hop scenario. More explicitly, RLNC provides a nature way to exploit multiple paths from the source to the destinations. This is because that coded packets are mixture of original packets. All relay nodes are free to send several coded packets, which will all be useful for decoding at the sink node. The CodeCast protocol proposed based on RLNC in [11] demonstrates benefits in terms of throughput and robustness improvement. By contrast, without network coding, relay nodes must be careful to avoid sending redundant packets, which results in a more complicated and less efficient routing protocol.

IV. NETWORK CODING ASSISTED SECURITY MECHANISMS

As shown in last section, network coding can be applied on video transmission for higher throughput. It can also be used in synergy with security mechanism [5] [12] [13]. In this section, we explain network coding assisted security mechanism in two aspects: video data encryption and multicast group key agreement.

A. Lightweight encryption algorithm

WoNC: In the absent of network coding, the data payload must be encrypted in order to protect data confidentiality against eavesdropping attack. Assuming that only the legitimate sink nodes have the decryption key, any intermediate node is not able to interpolate the video content based on the received encrypted packets. A comprehensive encryption methods for H.264 video can be found in [4].

RLNC: Network coded packets using RLNC are linear combinations of several original packets. If an intermediate node is not able to decode network code packets, it is not capable of interpolating the video content, which provides some degree of security. It also provides a lightweight solution as suggested in [12]. In this scheme, the network coding coefficients are encrypted using a secrete key at source node, then the encrypted coefficients becomes a part of the payload. Normal RLNC is further applied at relay nodes. The sink nodes who have the secrete key will first decode the coded packets using Gaussian elimination, then decrypt the source-node coding coefficients using the secrete key, and finally obtain the original video data by using Gaussian elimination again. The video content is protected at relay nodes who do not have the knowledge if the encryption key. Since the coding coefficient is significantly less than the payload, this mechanism significantly reduce the amount of encryption data.

We illustrate the above-mentioned encryption methods in Figure 2. Due to the adjustment explained in Section 6-B, RLNC is not applied on layer-1 video packets. Hence, the reduction using the lightweight encryption algorithm of [12] can only be achieved for enhancement layer video packets. In some use-cases, enhancement layer video packets also need to be encrypted. For example, in video-on-demand services, higher video quality is offer only to those users subscribe to this service.

B. *Efficient multicast group key agreement*

Aided with a trustworthy cluster-head as well as the connected authentication server (if available), it is possible to set up pair-wise keys between each cluster member node and the cluster head. We have explained this in Section II. However, for video multicast, a common group key is required between source node and all sink nodes. The management of group key is a difficult problem. Each time a new member is added or an old member is evicted from the group, the group key must be changed to ensure backward and forward security, which means a new member cannot figure out any past group key, and an evicted member is not able to guess any new group key neither.

Many algorithms have been proposed for group key management [8]. One of them is called One-way function tree (OFT) assuming the presence of a centralized trust framework. Taking the example scenario shown in Figure 1, we illustrates the conventional OFT algorithm in Figure 3. More explicitly, the cluster-head Node-R2 is serves as a central trust point. At the beginning, Node-S, Node-A and Node-B have an individual symmetric key with the cluster-head Node-R2 denoted as Ks, KA and KB, respectively. Then, these group members are divided into subgroups of size 2, and a new shared key between these two members is derived. For example, a new shared key between Node-S and Node-A is calculated as

$$K_{SA} = f\left(h(k_A), h(K_S)\right) \qquad (1)$$

where h(.) is a one-way hash function, and f (.) is a mixing function such as concatenation. These new shared keys are further divided into subgroups of size 2, and a higher-level shared key is agreed using the same method. This process continues until a single shared key is agreed among all member nodes. Since the cluster head have all the individual keys, this shared key could be calculated at cluster head. However, for better security, it is not a good idea to directly transmit this group key to each member nodes. Instead, the cluster head encrypts the one-way hash value using individual secret keys, and distribute them to group member nodes, so that the group key could be derived locally at each member node. For example, the cluster head transmits two encrypted message to Node-A E_(K_A) (h(K_S)) and E_(K_A) (h(K_B)) where E_(K_A) (.) represents encrypting a message using Node-A’s symmetric key KA .Similarly, the cluster-head must transmit 2 encrypted messages to Node-B and Node-S, which results in 6 messages in total. The reason of having a tree structure is to facilitate updating or regeneration of new keys when the group membership changes [8].

The above-mentioned communication overhead for group key initialization could be reduced by using a simple XOR network coding operation. Instead of transmitted two messages E_(K_A) (h(K_S)) and E_(K_A) (h(K_B)) to Node-A and Node-B respectively, the cluster-head could simply broadcast one message XOR (h(K_A),h(K_S)) to both Node-A and Node-B. Since they already have the hash value of its own key, it is easy to obtain the hashed value of the other key using XOR. Although the message is not encrypted, it is secure since only Node-A has the knowledge of its own key. As a result, this XOR-aided OFT algorithm could reduce 50% of the communication overhead in group key initialization process.

V. SIMULATION RESULTS

A. *Simulation Setup*

An end-to-end test-bed platform has been used for the experiments comprising video encoder, streamer, network emulator and decoder. The test-bed has been used to evaluate video quality of two video sequences. The video encoder uses an H.264/SVC encoder configured to create two Medium Grain Scalability (MGS) layers of one base and one enhancement layer. More specifically, two video sequences namely Crew and Soccer with 4-CIF (704x576) resolution has been used during the experimentation phase. The frame rate is 60 fps, and there are 600 frames lasting for 10 seconds. The intra period is set to 8 frames. The quantization parameters are set to 36-30 for both layers.

An H.264/SVC Streamer/Receiver has been implemented in order to transmit and receive each video sequence through the network using RTP/UDP/IP protocol stack. Each generated packet of the streamer has a single Network Abstraction Layer Unit (NALU) with a total size of 1400 bytes. Also an additional path is responsible for the transmission of the Parameter Sets (PS) to the client through a TCP/IP connection for more reliability because of their importance.

Between the streamer and the client, we consider the two-hop network shown in Figure 1. Video packets are generated at Node-S, relayed by Node-R1, Node-R2 and Node-R3. Node-A and Node-B are destination nodes. The whole experiment has been repeated quite a few times in order to be statistically correct.

B. *Simulation Results*

Figure 4 illustrates the PSNR versus time for both RLNC and WoNC schemes. For low packet loss rates, RLNC provides better robustness as opposed to WoNC in terms of PSNR variations. As packet loss rate increases above 5% the two schemes starting to converge.

Table 1 compares the amount of data needs to be encrypted using WoNC and RLNC scheme propose in Section VI. B and Section V.A. In both schemes, the payload in base-layer video packets needs to be encrypted. For enhancement-layer video packets, WoNC scheme encrypts all the payload, while RLNC scheme only encrypts the coding coefficients. We have set Galois Filed size equal to 8 2 , and variable generation size with three options 2, 3 and 4.

For both video sequence Crew and Soccer, about 50% reduction is achieved using RLNC scheme.

TABLE I: AMOUNT OF ENCRYPTION DATA FOR TWO VIDEO SEQUENCES USING 1) WONC SCHEME AND 2) RLNC SCHEME, AS WELL AS THEIR RATIO.

		WoNC (Bytes)	**RLNC (Bytes)**	**(RLNC/WoNC) %**
Crew	Layer-1	3823027	3823027	100%
	Layer -2	3477276	3419	0.09%
	Total	7300303	3826446	52.41%
Soccer	Layer-1	3757378	3757378	100%
	Layer-2	3226585	3418	0.1%
	Total	6983963	3760796	53.8%

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we have explored the idea of using network coding for video multicast over wireless ad-hoc networks. We have examined three schemes: network coding aided video transmission, network coding aided encryption, and XOR-aided one-way-function tree group key initialization. The simulation results using two video sequences have proved that network coding significantly improves the video quality because of its capability of utilizing multi-paths and retransmission efficiency. Using network coding also reduces about 50% of the amount of data for encryption, as well as 50% of the communication overhead in group key initialization. We could conclude that network coding is a promising energy efficient solution for video multicast in future network. In summary, using network coding for secure video multicast over wireless ad-hoc network is a promising technical, and worth further in-depth investigation.

ACKNOWLEDGMENT

The research leading to the present results has been undertaken within the European Union's Horizon 2020 research and innovation programme under grant agreement H2020-MSCA-ITN-2016 SECRET-722424

REFERENCES

[1] H. Schwarz, D. Marpe and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard", IEEE Trans. On Circuits and Systems for Video Technology, Vol. 17, No. 9, September 2007

[2] E. Magli, M. Wang, P. Frossard, and A. Markopoulou, "Network Coding Meets Multimedia: a Review," IEEE Trans. Multimed., vol. 15, no. 5, pp. 1195 – 1212, 2012.

[3] H. Seferoglu and A. Markopoulou, 'Video-Aware Opportunistic Network Coding over Wireless Networks", IEEE JSAC, Vol. 27, No.5, pp. 1-16, 2009

[4] T. St and A. Uhl, "A Survey of H . 264 AVC / SVC Encryption," Technical report, 2010.

[5] L. Lima, S. Gheorghiu, J. Barros, M. Medard, and A. L. Toledo, "Secure network coding for multi-resolution wireless video streaming," IEEE J. Sel. Areas Commun., vol. 28, no. 3, pp. 377–388, Apr.2010

[6] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network Information Flow," IEEE Trans. Inf. Theory,, vol. 46, pp. 204–1216, 2000

[7] Ning Cai and T. Chan, "Theory of Secure Network Coding," Proc. IEEE, vol. 99, no. 3, pp. 421–437, Mar. 2011.

[8] B. Wu, J. Wu, and M. Cardei, "Chapter 30: A Survey of Key Management in Mobile Ad Hoc Networks," in in Handbook of research on wireless security, 2008, pp. 176–188

[9] T. Matsuda, T. Noguchi, and T. Takine, "Survey of Network Coding and Its Applications," IEICE Trans. Commun., vol. E94-B, no. 3, pp. 698–717, 2011.

[10] H. Wang, J. Liang, and C. Kuo, "Overview of robust video streaming with network coding," J. Inf. Hiding Multimed. Signal Process., vol. 1, no. 1, pp. 36–50, 2010

[11] J. Park, M. Gerla, and D. Lun, "Codecast: a network-coding-based ad hoc multicast protocol," IEEE Wirel. Commun., vol. 13, no. 5, pp. 76–81, 2006

[12] J. Vilela, L. Lima, and J. Barros, "Lightweight Security for Network Coding," in IEEE International Conference on Communications (ICC '08), 2008, pp. 1750 - 1754

[13] R. Zeng, Y. Jiang, C. Lin, Y. Fan, and X. (Sherman) Shen, "A scalable and robust key pre-distribution scheme with network coding for sensor data storage," Comput. Networks, vol. 55, no. 10, pp. 2534–2544, Jul. 2011

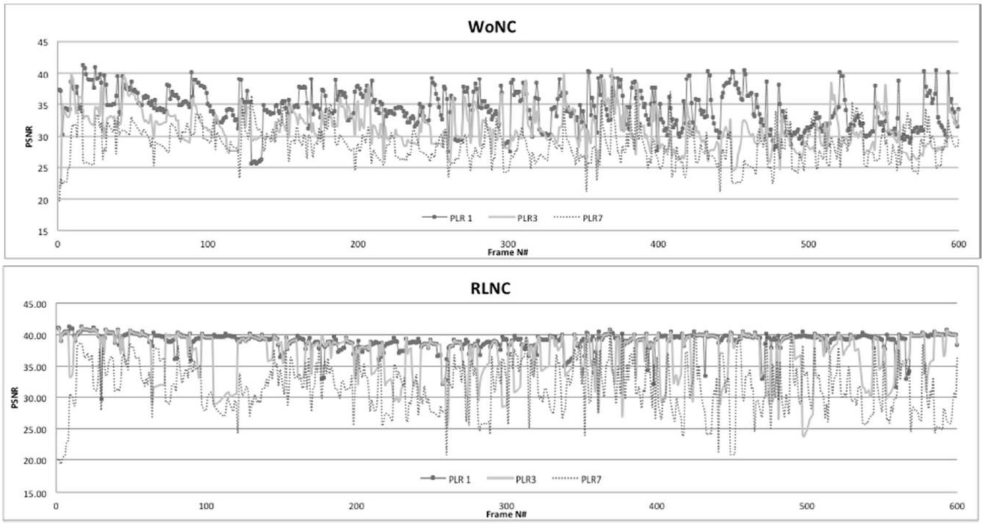

Figure 3. PSNR versus time for both RLNC and WoNC for various packet loss rates (1%, 3%,7%).