Computer Communications 34 (2011) 1661–1672

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Review

H.264 video transmissions over wireless networks: Challenges and solutions

Yi-Mao Hsiao [a], Jeng-Farn Lee [b,*], Jai-Shiarng Chen [a], Yuan-Sun Chu [a]

[a] Department of Electrical Engineering, National Chung Cheng University, Chia-Yi, Taiwan
[b] Department of Computer Science and Information Engineering, National Chung Cheng University, Chia-Yi, Taiwan

ARTICLE INFO

Article history:
Received 16 August 2010
Received in revised form 31 January 2011
Accepted 26 March 2011
Available online 2 April 2011

Keywords:
H.264
SVC
Cross-layer design
Wireless networks

ABSTRACT

Multimedia video streaming is becoming increasingly popular. Using multimedia services, there are more and more users in end-system over wireless networking environment. H.264/AVC is now the standard for video streaming because of its high compression efficiency, robustness against errors and network-friendly features. However, providing the desired quality of service or improving the transmission efficiency for H.264 video transmissions over wireless networks present numbers of challenges. In this paper, we consider those challenges and survey existing mechanisms based on the protocol layers they work on. Finally, we address some open research issues concerning for H.264 video transmission in wireless networks.

1. Introduction

IEEE 802.11 and 802.16-based wireless networks are promising technologies for accessing the Internet due to the characteristics of low cost, robustness, and ease of deployment. Moreover, various wireless applications are being developed due to the improved display capacity and execution power of mobile devices, e.g., PDAs, notebooks, and dual-mode handsets. Several video streaming applications, such as wireless internet protocol television, Radio over IP, and Video conferencing, have been developed. For video streaming, H.264/AVC has become the standard for video streaming due to its high compression efficiency, error robustness and network-friendly features. Previous works on wired networks focus on the resource reservation among network devices [1,2], providing differentiated services [3–5] and the codec enhancements [6,7]. However, due to the characteristics of wireless networks such as heterogeneous wireless uses, high error rate, limited bandwidth, multiple transmission rates, time-varying channel conditions and dynamic network users, providing QoS for wireless video streaming presents several new challenges. Although IEEE 802.11e Enhanced Distributed Channel Access (EDCA) provides a priority-based scheme for different access categories (ACs), but it relies heavily on user experience to configure the EDCA parameter sets properly. IEEE 802.16 WiMAX provides the rtPS service class for video streaming, but resource allocation among the multiple relay base station (MR-BS) and relay stations (RSs) for multicast or broadcast video streaming service in wireless relay networks is a new research area. Previous approaches for the resource allocation problems work for serving all multicast members with minimal resource/energy in the wireless ad hoc networks where all nodes can serve as a relay for the other nodes [8–10]. However, only RSs can relay data in WiMAX relay networks. In addition, another maximization problem to serve maximal recipient within a resource budget needs to be addressed due to rare wireless resource or call admission control policy. Finally, how to provide high performance end-user systems over wireless networks is a hot research issue until now since the capacity limitation of mobile devices and the required computing power for video streaming is high. This research issue focuses on the efficiency of decoder and TCP/IP implementations on mobile devices [11–15]. Therefore, we still need to address the issues about providing the desired QoS or improving the transmission efficiency for H.264 video transmissions over wireless networks. The video transmissions discussed in this work focus on IP Unicast. The Multicast services in WiMAX relay networks are also achieved by multiple IP Unicast via the MR-BS and RSs, and the broadcast nature of wireless media. Besides, researches for both non-scalable [3,4,14–22] and scalable [1,5–7,23–29] video transmissions are covered.

The remainder of this paper is organized as follows: Section 2 provides an overview of H.264/AVC. In Section 3, we discuss the challenges of the provision of QoS for H.264/AVC and resource allocation in wireless networks. In Section 4, we classify existing mechanisms according to their designed protocol stacks; and in Section 5, we consider some open research issues concerning H.264 video transmissions over wireless networks. Section 6 contains some concluding remarks.

* Corresponding author. Tel.: +886 5 2720411x33128; fax: +886 5 2720859.
E-mail address: jflee@cs.ccu.edu.tw (Y.-M. Hsiao).

0140-3664/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.comcom.2011.03.016

2. Overview of H.264/AVC

In March 2003, the Video Coding Experts Group (VCEG) and the Moving Picture Experts Group (MPEG) formed a Joint Video Team (JVT) to finalize a new video coding standard called H.264/AVC. The standard covers two layers: the Video Coding Layer (VCL), which creates a coded representation of the source content, and the Network Abstraction Layer (NAL) to format the VCL data and provide header information about how to use the data for video delivering over network [16].

2.1. Video coding layer (VCL)

The VCL in H.264/AVC provides improved flexibility and adaptability in video transmission. An image is partitioned into smaller coding units called *macroblocks*, which are comprised of slices that can be parsed independently of other slices in the picture. The VCL layer partitions slices into three groups: (i) Partition A defines macroblock types, quantization parameters, and motion vectors; (ii) Partition B is the intra partition; and (iii) Partition C is the inter partition. The slice groups allow flexible partitioning of a picture into slices. These partitions are based on a slice group map that is specified by the content of the picture parameter set and header information. The slice header information assigns a unique slice group identifier to each macroblock.

2.2. Network abstraction layer (NAL)

NAL units are the video data encoded by VCL and one-byte header that shows the type of data in the NAL unit. One or more NAL units can be encapsulated in a transport packet. The encoded video data in NAL units is classified into (1) VCL NAL units, which are coded slices or coded slice data partitions and (2) non-VCL NAL units, which contain associated information, such as the sets of parameters and supplemental enhancement information (SEI). The SEI stores the introductions, copyright and user definition of a video stream. A coded video sequence represents an independently decodable part of a NAL unit bit stream. The sequence starts with an instantaneous decoding refresh (IDR) access unit. The IDR access unit and all following access units can be decoded without decoding any previous pictures of the bit stream. The Nal_Ref_Idc (NRI) of an NAL unit header contains two bits that indicate the transmission priority of the NAL payload. The 11 (Parameter Set Concept) of NRI is the highest priority, followed by 10 (Coded Slice data partition A), 01 (Coded Slice data partition B) and 00 (Coded Slice data partition C), which is the lowest priority. For Parameter Set Concept (PSC) contains information such as picture size, optional coding modes employed, and macroblock to slice group map. To provide the desired QoS, the information in NRI can be referenced as the importance of the packets when the video data is transmitted.

3. Challenges of video transmission over wireless networks

In this section, we consider the challenges of H.264/AVC video transmission over wireless networks.

3.1. Unnecessary retransmissions

In a wireless network, every mobile host connects to the Internet via an access point, a base station or a relay station. The video frames will be retransmitted when transmission errors or collisions occur during frame delivery. However, unnecessary retransmissions may be performed if the video frame has missed its play-back time, resulting in bandwidth waste and further channel access contention. Moreover, in wireless network MAC protocols, if the frame checksum is failed, the frame will be dropped at the receiver side, even if the whole packet is received. This also results in bandwidth waste, since these error frames can still be used by bit error resilient mechanisms of the MAC and upper layer protocols.

3.2. Bandwidth fluctuations

In a wireless network, data transmission is via the wireless radio medium. Node mobility causes bandwidth fluctuations due to differences in channel quality. Even if the wireless station is stationary, its wireless bandwidth may fluctuate due to multi-path fading, co-channel interference, and noise disturbances. Bandwidth fluctuations represent the main challenges of real-time video streaming over wireless networks. For this reason, a cross-layer design is needed for video streaming in wireless networks. The application layer should encode the video streaming according to the available bandwidth based on the physical layer and the contention parameters of the MAC layer. The MAC layer should transmit or drop frames according to their importance or priority based on the context of the video frames (e.g., the NRI in NAL unit). Moreover, the parameter sets or bandwidth assignment policy of the MAC layer should be adjusted dynamically based on the number of video stations, as well as the requirements and channel qualities of individual receivers.

3.3. Contention-based MAC protocol

Although the Point Coordination Function (PCF) of IEEE 802.11 and HCF Controlled Channel Access (HCCA) of 802.11e provide polling-based services for real-time frame delivery, they are seldom implemented in off-the-shelf WiFi devices due to the implementation complexity and strong assumption of global synchronization. Both 802.11 DCF and 802.11e EDCA are contention-based MAC protocols, so it is difficult to provide guaranteed service under DCF and EDCA. Although 802.11e EDCA provides relative differentiated service among different access categories (ACs), how to map H.264/AVC frames to the same or different EDCA ACs is a major research issue. Moreover, the performance of high priority ACs in the original EDCA (i.e., the parameter sets of EDCA suggested in the 802.11e standard) is very sensitive to the number of active low priority flows. Thus, how to adjust the EDCA parameter sets dynamically based on the context of video streaming or the network environment to provide efficient and effective video delivery in wireless networks is a challenging issue.

3.4. Heterogeneous wireless users

Wireless users may have different requirements or utility functions for the same video stream because of their differences in the display capacity, processing power, and battery life of end-system equipment. They may also experience different channel conditions such that the source or relay nodes need different amounts of resources (e.g., transmission time, power or bandwidth) to deliver the same data to different users. In addition, different allocations of resources to the source or relay nodes change the network topology. If a sender node (i.e., the source node or relay nodes) transmits the video streaming with different transmission power or modulation codes, the nodes that can successfully receive the streaming data change. Consequently, the network topology and thus the routes from the SSs to the source node are also changed. As a result, improving the efficiency of the transmission scheme, scheduling and resource allocation, relay node selection and path construction in wireless relay networks for video streaming multicast programs still requires a great deal of research effort.

4. Existing solutions

A number of approaches have been proposed to improve video streaming transmission over wireless networks. We classify these approaches based on the network protocol stacks they work on and the system on chip (SoC) design. The involved protocol stacks are the application layer, the transport layer, the network layer and cross-layer (i.e., the designed mechanisms involve operations from more than one protocol stack). The main design issues are (1) how to improve the terminal capability of mobile users; (2) how to minimize the overall impact of frame losses in wireless networks and (3) how to improve the transmission efficiency and resource utilization for video streaming delivery. In the application layer, the major issue is the coding efficiency of the codec. Scalable Video Coding (SVC) has been proposed to provide scalability of video coding. In the transport layer, bit-error resilience is a technique that passes packets with bit errors to the higher layer instead of dropping them. With regard to the network layer, the main issues are path construction, relay node selection and resource allocation to deliver H.264/AVC streams to groups of users in wireless relay networks with the minimal resources or to serve maximal users with a limited resource budget. In the cross-layer design, information from the application layer, the MAC layer and the physical layers is considered together in order to improve video delivery performance. Finally, the terminal capability is also an important issue in wireless communications since video decoding itself costs a heavy computing power. Advance hardware architecture of SoC design is proposed with very large scale integration (VLSI) technology to improve video delivery to end users.

4.1. Scalable video coding

Scalable video coding [7] has been proposed as an extension of the H.264/AVC standard. The objective of SVC is to encode a video stream with one or more subset bit streams, which can be decoded with a complexity and reconstruction quality similar to that of the data encoded by the H.264/AVC design. SVC is flexible and adaptable because it only needs to encode a video once and the resulting bit stream can be decoded at multiple reduced rates and resolutions. SVC provides three types of scalability: temporal scalability, spatial scalability and SNR (Quality) scalability. For temporal scalability, H.264/SVC likes H.264/AVC that uses hierarchical B picture to do motion compression as shown in Fig. 1. Decoding the base layer provides low but standard video quality, and decoding the base layer with the enhancement layers improves the quality of video streaming. In Fig. 1, except I-frames, every frame in the group of picture has to refer to neighbor frames. For example, the second B3-frame needs the first B2-frame and the B1-frame to decode. The reference relationship of different frames can be classified into four temporal layers. The base layer has the most important role since if the I/P-frames are lost due to transmission errors or collisions, all frames would not be decoded. The frames on higher layers are less important since less frames are decoded based on them. Fig. 2 shows a multi-layer structure with additional inter-layer prediction for spatial scalability of SVC where the video content is encoded into several resolutions and frame rates. When the frames are transmitted over the Internet, the media gateway with a filter can patch suitable resolutions and frame rates to media clients based on their network conditions. SNR scalability supports fine grain scalability (FGS) which is composed of base layer and enhancement layer. The base layer provides basic quality and the enhancement layer called progressive refinement (PR) slice represents a refinement to the residual signal and can be truncated at any arbitrary point as shown in Fig. 3. Each PR slice needs to refer to the corresponding slice in the base layer. The base layer uses nonscalable coding which is more efficient in terms of compression ratio than scalable coding used in the enhancement layer. The size of SVC NAL unit header is four bytes. The second byte contains Reserved Bits and Simple Priority and third byte represents Temporal, Spatial and Quality level. These fields can be referenced as the importance of the packets when the video data is transmitted.

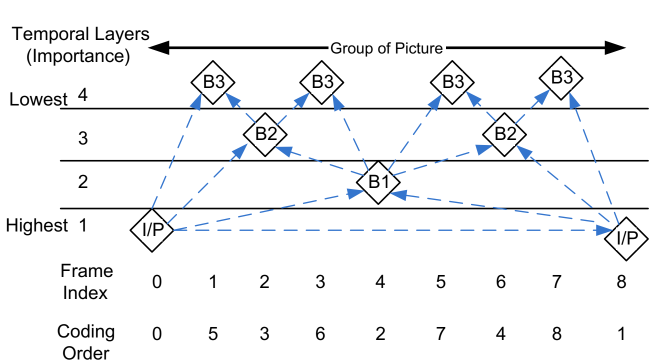

Fig. 1. An illustration of temporal scalability of SVC.

The H.264/SVC video streams are much vulnerable to transmission errors and packet losses. The effects not only corrupt the current frame, but also propagate to subsequence frames. Thus, the end users experience non-graceful performance degradation. Refs. [23–25] show the non-graceful performance degradation of H.264/SVC video for different packet loss ratios. Therefore, this calls for mechanisms to protect video frames with different importance or priorities. Ref. [23] addresses the problem of unequal error protection (UEP) for scalable video transmission over wireless networks. Unequal amounts of protection data should be allocated to different bit-stream of video streaming to provide a graceful degradation caused by packet losses. Ref. [23] uses a genetic algorithm (GA) to quickly get the allocation pattern. Schierl et al. [24] figure out that the robustness of a streaming connection against packet losses can be significantly increased if the different layers of the coded video streaming are unequally protected by a forward error correction scheme. Jang et al. [25] propose adaptation mechanisms based on Access Unit (AU) or GOP of SVC video streaming to reduce the error propagation caused by packet losses by selectively discarding less important frames with the layer-dependency consideration.

4.2. Bit-error resilient

Bit-errors in the radio channel reduce link utilization in wireless networks, especially for real-time video streaming services. Even if there is just one bit-error, the receiving packet has to be dropped. Larzon et al. propose UDP-Lite [17] to allow delivery of corrupted packets using a lightweight checksum calculation. Traditional UDP uses full checksum, which is comprised the header and payload of a packet. In UDP-Lite, Packets are divided into sensitive and insensitive parts, and the checksum is calculated only based on the sensitive part. UDP-Lite uses the length field in the UDP packet header to indicate the length of the sensitive part. Thus, the bit- error bits may not be in the checksum coverage (i.e., the sensitive part), so packets are passed to the application layer, even if they contain bit errors. Korhonen et al. propose packet bit-error resilient strategies for H.264/AVC video streaming over wireless networks [18]. The proposed mechanism uses small slices and protects the most vulnerable bits by using the UDP-Lite protocol as shown in Fig. 4. The NAL units are split in two parts. The first part contains the most relevant bytes, including the slice header, and the second part is the macroblock. The protected part of each slice is allocated in the beginning of the packet payload, preceded by the length information of the slice. The macroblock data resides in the unprotected part. Shorter protected portion of UDP-Lite packets is

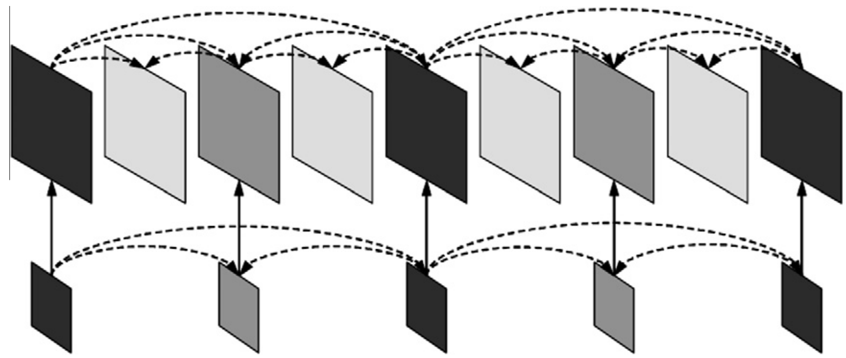

Fig. 2. An illustration of spatial scalability of SVC.

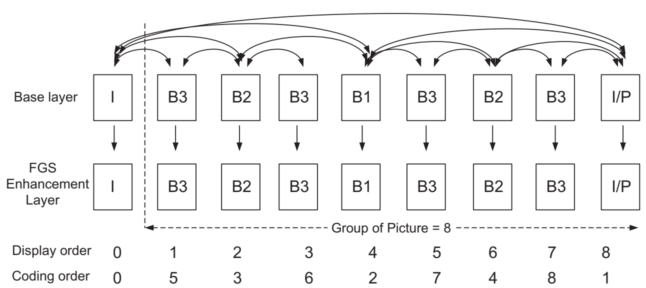

Fig. 3. SNR scalability.

used to reduce packet losses in the transport layer. They simulate Foreman and Soccer H.264 sequences with three kinds of error model: random errors, short bursts and long bursts. The Foreman and Soccer video are with 512 kb/s and 256 kb/s. The simulation results show that the most improvement of peak signal to noise (PSNR) is 14.72% in random error model with packet size 150 bytes for Foreman video with 256 kb/s. The improvements of PSNR in short bursts and long bursts error models with packet size 450 byte for Foreman video with 512 kb/s are 9.6% and 7.8%, respectively.

4.3. Path construction or relay selection for multicast/broadcast services in wireless relay networks

In a wireless relay network, a new class of infrastructure node called a Relay Station (RS) is used to improve the performance and coverage of the network. Since each mobile station (MS) may have heterogeneous channel conditions, the amount of resources required to receive streaming frames successfully from the source node or RSs is different. Since the resource allocated to RSs (i.e., transmission time or power) affects the topology of the wireless relay network, it also influences the path constructed from each MS to the source node to receive the streaming data. In such a relay network environment, how to allocate resources to the multiple relay base station (MR-BS) and the RSs to maximize the number of MSs that can be served given a resource budget is a challenging issue.

For example, consider a single-level (i.e., the RSs are only allowed to relay data for MSs, but not for other RSs) WiMAX multi-hop relay networks shown in Fig. 5, in which there are one MR-BS, 12 RSs, and 17 MSs. Suppose that the multicast program is allocated 22 units of the resource. Our objective is to allocate the resource among the MR-BS and RSs such that as many as MSs can be served (i.e., receive the video streaming successfully). Different types of wireless resource are utilized in wireless relay networks. Without loss of generality, the resource here refers to the amount of the transmission medium that can be distributed and utilized by different nodes. Therefore, depending on the physical design of the wireless network, the resource can be the number of timeslots, sub-channels or the transmission power. For example, in a WiMAX relay network, the channel resource can be the total time slots in a TDD super frame spent for a multicast IPTV program. The required time slots for transmitting a multicast stream differs as MSs and RSs have different channel conditions, resulting in different modulation schemes and thus transmission rates required to successfully receiving data. In the following examples, we use a well known channel model [30] to determine the channel quality based on the node distribution. In this model, the required resource is

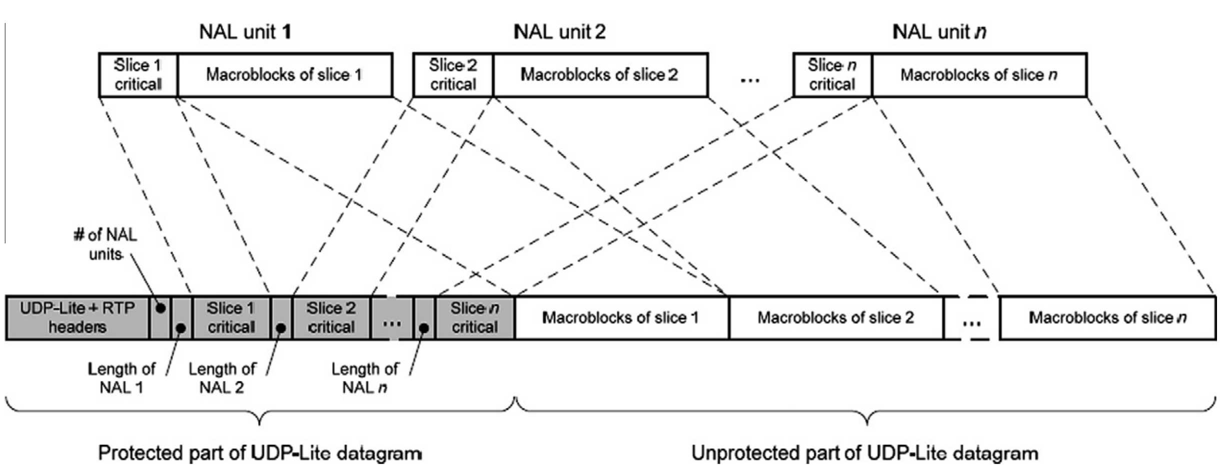

Fig. 4. Bit error resilient packetization scheme for UDP Lite in [18].

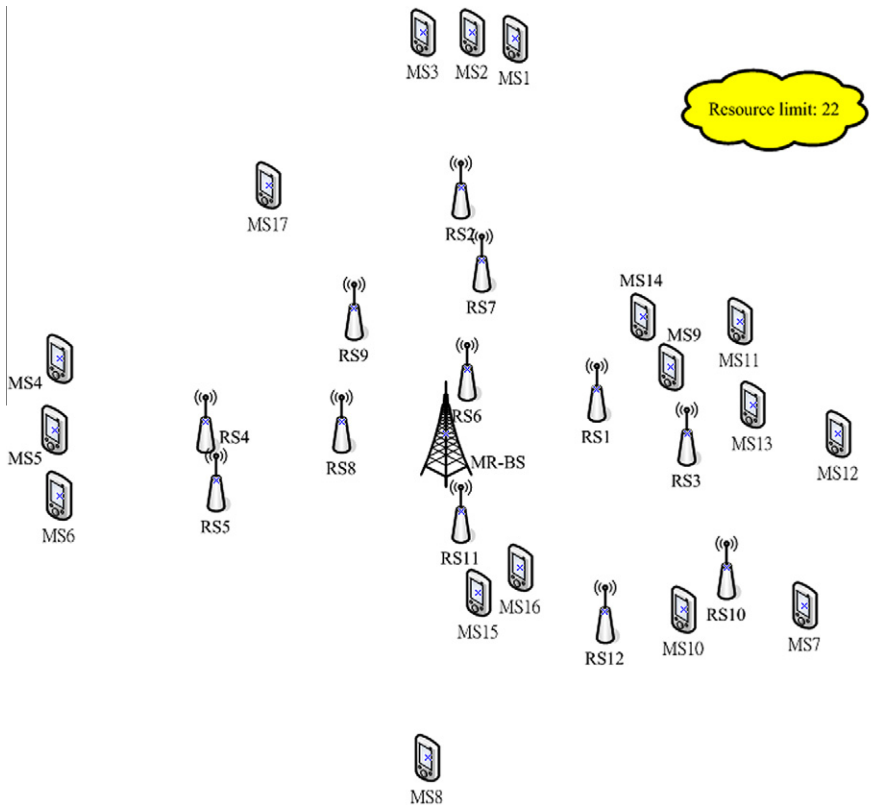

Fig. 5. An illustration of WiMAX relay network with resource budget 22.

represented as $\frac{1}{d^a}$, where d is the distance between the sender and receiver and a is the channel attenuation factor that is 2 in this work.

If we do not consider any RS relaying from the MR-BS, then the solution to the resource allocation problem is trivial, but not efficient. As shown in Fig. 6, we simply employ all available resource to the MR-BS (i.e., the MR-BS is allocated resource 22), and at most 8 MSs could be directly served by the MR-BS. Nevertheless, the problem immediately becomes tougher as we take RSs into account since different resource allocations among the MR-BS and RSs lead to different network topologies. One of intuitive solutions is to make each MS be served by the closest RS. Once the relay of each MS is determined, the path from each MS to the MR-BS can also be decided. Then, we can calculate the utility of serving a MS as the number of served MSs divided by the extra resource to serve that MS. We allocate the resource based on the decreasing order of the utilities of MSs until the residual resource is insufficient to serve any more MSs or all MSs are served. Fig. 7 demonstrates such an allocation where the MR-BS, RS_3, RS_4, RS_5, and RS_{11} are allocated 9, 4, 4, and 1 unit of resource, respectively. As a result, RS_3, RS_4, RS_5, and RS_{11} are being relay nodes of the MR-BS such that MS_9, MS_{11}, MS_{12}, MS_{13}, and MS_{14} can be served by the RS_3; MS_4 and MS_5 are served by the RS_4; MS_6 are served by the RS_5, and finally MS_{15} and MS_{16} are served by the RS_{11}. Therefore, 10 MSs can be served in Fig. 7. Although the allocation strategy demonstrated in Fig. 7 is better than that of directly served by the MR-BS since more MSs can be served in Fig. 7, it is not the optimal solution. As shown in Fig. 8, the optimal solution via a brute-force manner actually could serve 14 MSs within the resource budget 22. By above examples, the resource allocation problem for multicast/broadcast receipt maximization over wireless multi-hop relay networks is really a challenge.

Kuo and Lee first show that the multicast recipient maximization (MRM) problem with resource budget problem is NP-hard and propose a utility-based polynomial-time algorithm to solve it in [31]. However, this work is based on the wireless relay networks where the path from each MS to the MR-BS is predetermined. Kuo and Lee in [32] then propose another heuristic algorithm to solve the path construction problem in more general wireless relay networks. The algorithm first assigns the entire budget to the BS, which means the solution does not utilize any RSs in this step. Then it tries to give some of the BS's resource to other RSs by choosing an MS as the farthest node that the BS can serve and releasing the remaining resources to certain RS step-by-step to improve the resource allocation. Performance evaluations via simulations show that the results of the proposed algorithm are very close to the optimal solutions under different network topologies and resource budgets.

4.4. Cross-layer design

Bandwidth fluctuations and time-varying wireless network conditions call for a cross-layer design that breaks the boundaries of traditional protocol stacks to improve the QoS of H.264 video streaming in wireless networks.

Fig. 9 shows the architecture and components of the general cross-layer design, which includes control schemes at the application, MAC and physical layers. The network condition estimator is used to estimate conditions like the bit-error rate and the available bandwidth of certain service class. Then, the models in the application layer can minimize the distortion of the video quality by optimal bit allocation, reducing the number of bits required for forward error correction, and determining the priorities of packets according to the impact of their loss. However, because of the slower timescale of the application layer, it is difficult for the layer to respond to rapid variations in bandwidth. Therefore, the component in the MAC sub-layer reacts by adjusting the transmission rate further, subject to minimum distortion. Then MAC sub-layer can map the

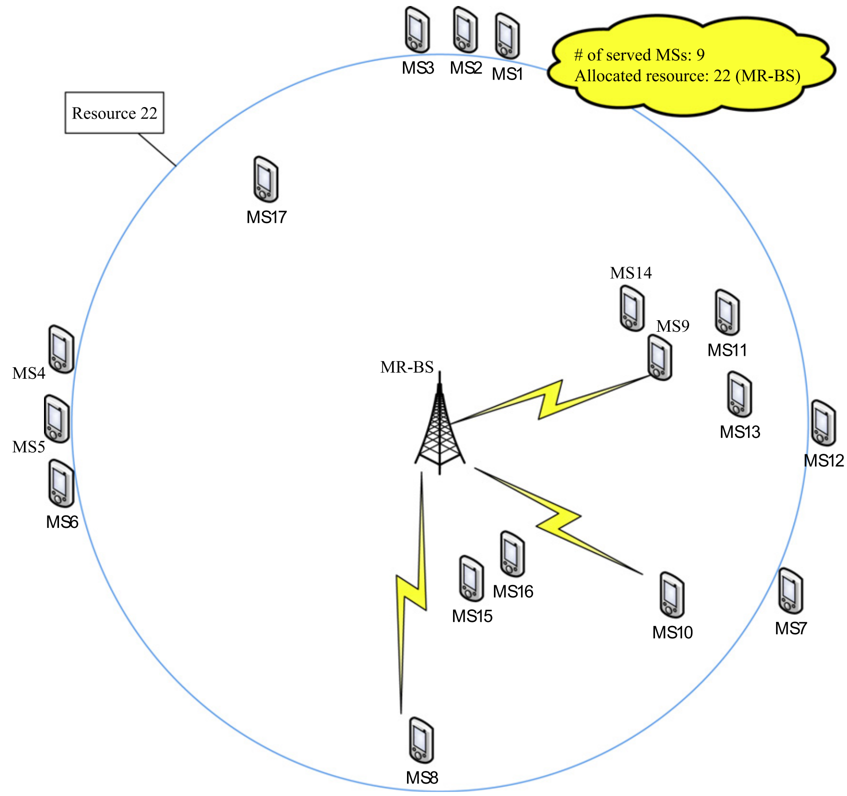

Fig. 6. At most 8 MSs could be directly served by the MR-BS when all budget is allocated to the MR-BS alone.

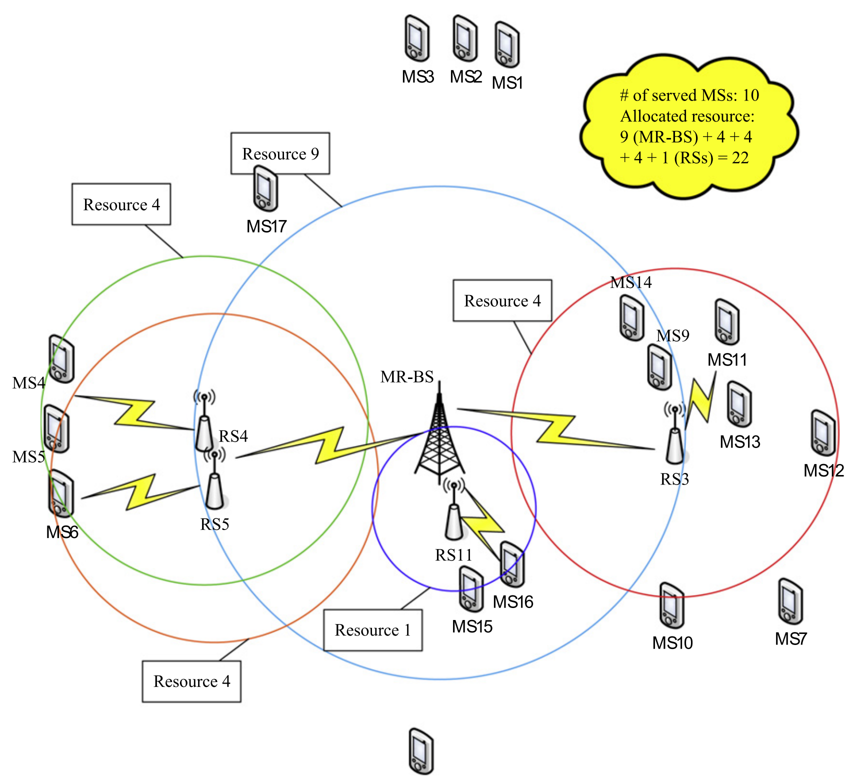

Fig. 7. 10 MSs can be served if each MS is served its closest RS.

streaming frames into one or more service classes and drops them based on their priorities. It also uses aggregation and fragmentation mechanisms to adjust the frame size, and adjust the transmission parameters, such as the retry limit, EDCA-parameter sets in IEEE

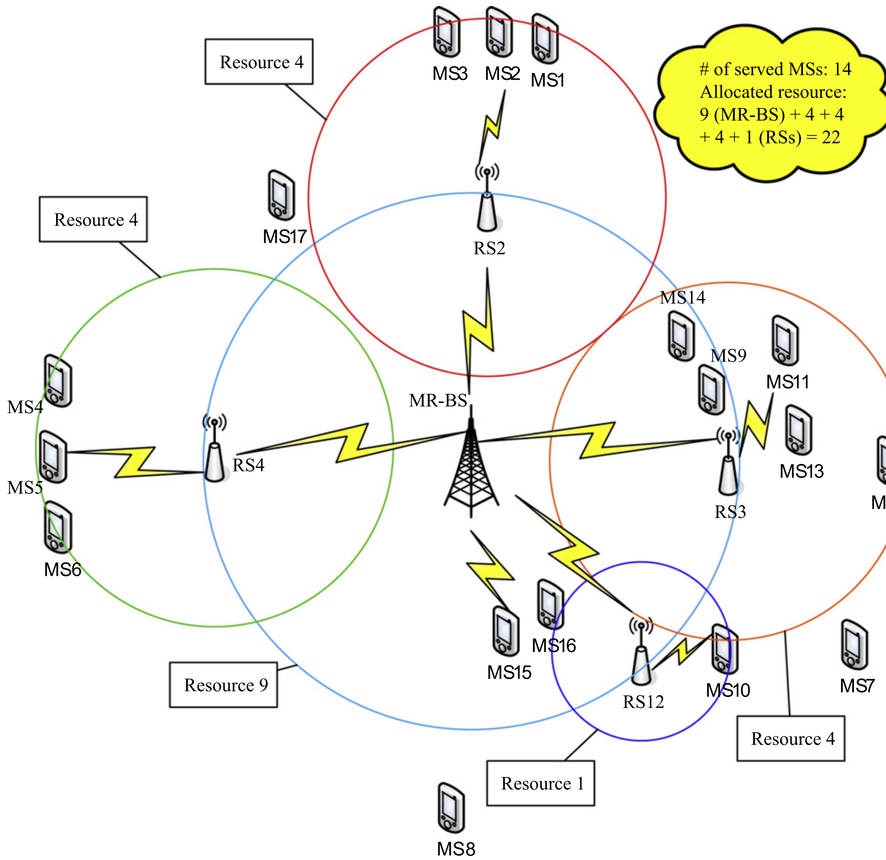

Fig. 8. The optimal resource allocation for the scenario in Fig. 5. 14 MSs can be served in the optimal solution.

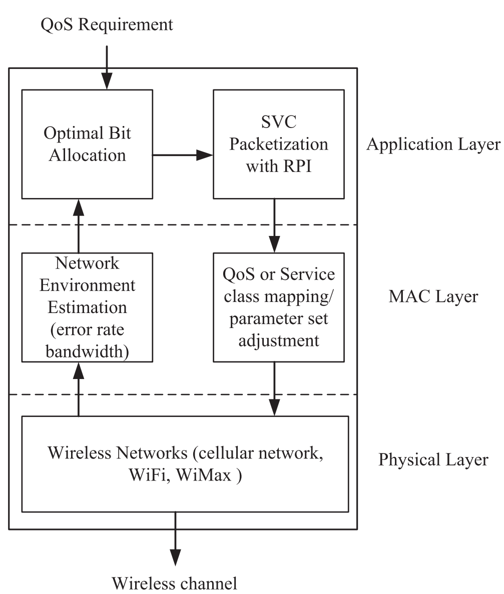

Fig. 9. The general architecture and components of the cross-layer design.

802.11e and the bandwidth assignment parameters in IEEE 802.16 WiMAX.

In [26], the authors propose the first cross-layer adaptive framework for video transmission over wireless networks, as shown in Fig. 10. The framework is comprised of three components: scalable video representation, network-aware end-systems and adaptive services. The scalable video representation uses a scalable video coding scheme. The network-aware end-system monitors the network's status (e.g., the bits-error rate and available bandwidth) and then adapts the video streams accordingly by adjusting the transmitted video representation at the application layer. The network-aware end-system can also drop packets at the MAC layer in a way that gracefully degrades the stream's quality instead of corrupting the flow outright caused by randomly dropping packets indiscriminately. Finally, the adaptive service modules provide adaptive QoS support for the scalable video during transmission such as reserving a minimum bandwidth to meet the demand of the base layer, adapting the enhancement layers based on the available bandwidth and the fairness policy. Consequently, the perceptual quality of video streaming changes gracefully during the fluctuation of wireless channel.

Many research approaches are proposed to improve the QoS for video streaming over IEEE 802.11 based wireless networks. Fig. 11 shows a classification of the cross-layer approaches. These QoS

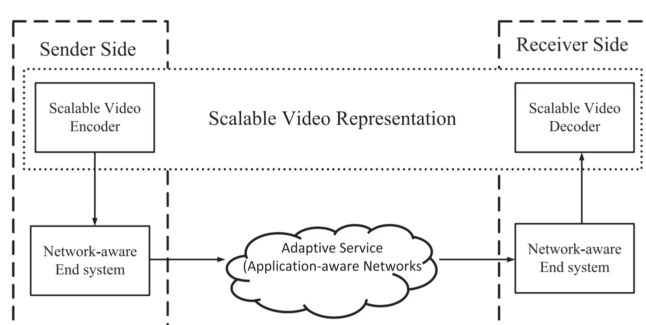

Fig. 10. An adaptive framework proposed in [26].

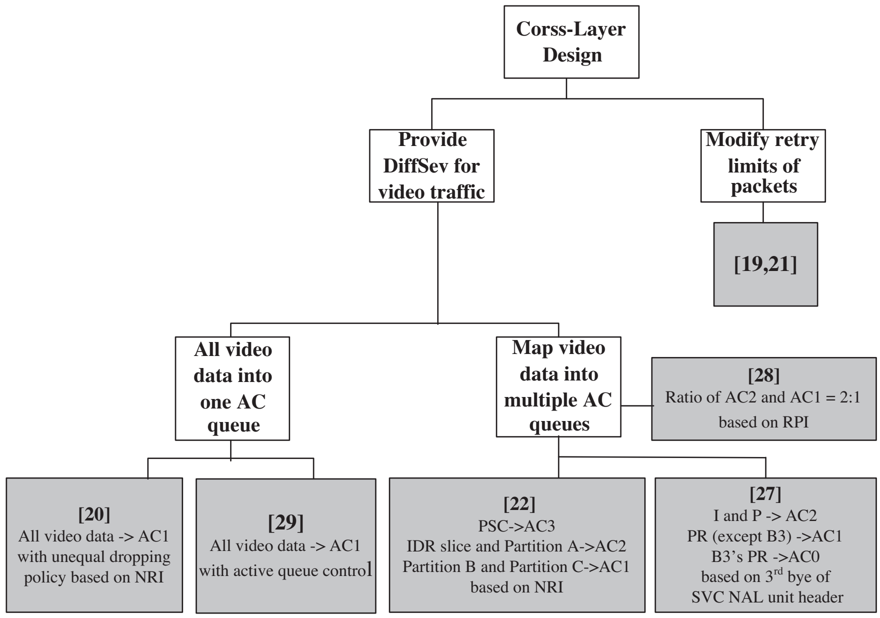

Fig. 11. The classification of the cross-layer approaches for IEEE 802.11 based wireless networks.

mechanisms provide desired QoS for video streaming traffic via adjusting the retry limits of video frames dynamically or provide differentiated service to different types of video frames by AC mapping or queue management mechanisms.

Transmissions over wireless networks require retransmissions to successfully deliver data to a receiver in case of packet losses due to transmission errors or collisions. However, retransmissions for video streaming traffic should consider the delay constraint and loss impacts of frames; otherwise, the wireless resource of the retransmission is waste. In [19,21], the authors propose retry limit adaptation schemes which dynamically adapts the retry limit based on the duration of retransmission deadline for each frame and its loss impact. The frameworks of the proposed mechanisms adopt a cross-layer scheme to calculate the loss impacts and estimated transmission times of packets at application-layer and MAC-layer, respectively. Then, they can dynamically determine whether to send or discard the video frames. Simulation results show that the proposed retry adaptation schemes significantly outperform the traditional static-retry-limit mechanism and the state-of-the-art time-based retry adaptation method in terms of packet loss and visual quality.

The QoS mechanisms of providing differentiated service based on importance of video frames can be classified into multiple ACs and single AC. The multiple ACs approaches put video frames into multiple ACs in 802.11e EDCA wireless networks. Ksentini et al. illustrate the idea of mapping different video packets into different EDCA ACs [22]. The authors address H.264 wireless video transmission over IEEE 802.11 Wireless LANs by proposing a robust cross layer architecture which exploits the IEEE 802.11e MAC protocol and the H.264 error resilience tools at application-layer, namely data partitioning. The cross layer architecture is shown in Fig. 12. By enabling data partitioning, the VCL layer divides the original video stream into several partitions (i.e., PSC, IDR pictures, Partition A (MB types, quantization parameters, and motion vector), Partition B (The Intra partition), and Partition C (The Inter partition)). They proposes a mapping algorithm that uses the NRI field, which in NAL header, to map the H.264 stream to a suitable AC. In the mapping algorithm, the most important information (i.e., PSC), are mapped onto the highest AC (i.e., AC [3] or AC_VO). In addition, IDR slice and Partition A are mapped onto to AC [2] (i.e., AC_VI) to guarantee a bounded delay and minimal loss rate. The Partition B and Partition C, which have less effect on QoS requirement, are mapped onto less priority access category AC [1] (i.e., AC_BK). The architecture increases the perceived video quality over that obtained by both DCF and EDCA. In [27], Chen et al. obtain six kinds of slices of SVC: I, P, B1, B2, B3 and PR which are the same as shown in Fig. 3. The slices of all frames at the base layer are transmitted with the AC_VI. The slices of enhancement layer PR (except B3) are mapped onto AC_BK. B3's PR slice has little effect to the playback quality such that they are mapped onto the lowest priority access category AC_BE. Foh et al. put the video streaming frames into AC_VI and AC_BK [28]. They find the saturation throughputs between AC_VI and AC_BK under different network load conditions by analyzing the service rates between AC_VI and AC_BK. Then they can determine the ratio of video frames put into AC_VI and AC_BK for the purpose of similar queue usage. On the purpose to provide adaptive queueing mapping in the network sublayer, it requires the application layer to provide the relative priorities of video packets. Relative Priority Index (RPI) is used to define the importance of different packets based on their loss impact. Although the video streaming frames of above approaches are put into more than one ACs to provide differentiated service or service protection for higher priority video frames, streaming frames in AC_BK and AC_BE must contend for wireless resource with other data traffic. When the traffic load of these data traffic is high, the performance of streaming traffic cannot be guaranteed.

In single AC design, all video streaming frames are put into AC_VI. Thus, the proposed QoS mechanisms use different drop probabilities for video frames or control the number of active stations transmitting video traffic to guarantee the desired QoS for video traffic. Chen et al. propose a cross-layer mechanism to improve the quality of H.264 video when encountering short-term bandwidth fluctuations over IEEE 802.11e wireless networks [20]. The mechanism consists of slice classification at application layer, dynamic packet selective transmission (DPST) at MAC layer, and channel condition prediction at physical layer. The DPST selects packets with highest priority to transmit so as to decrease the performance fluctuation of random drop. Zhang et al. control the number of "active" nodes on the channel to reduce collisions under intensive competition [29]. A distributed on–off queue control

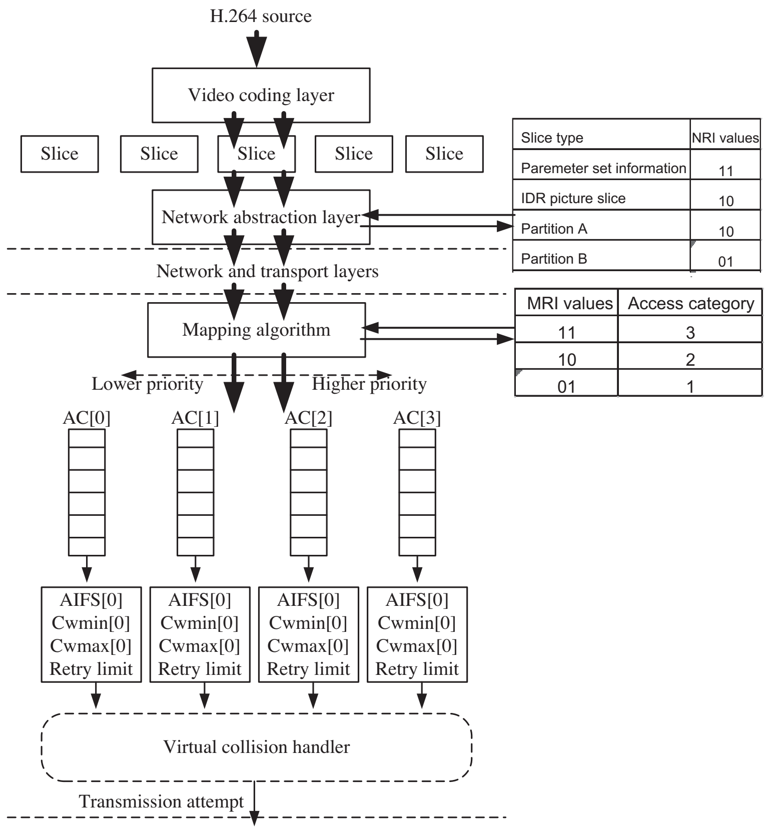

Fig. 12. The cross-layer QoS architecture proposed in [22].

mechanism (DOQC) adaptively maintains a certain number of "active" nodes transmitting the video traffic on the network so that the network can be operated at high throughput without going into congestion collapse. In addition, they use the low priority early drop (LPED) algorithm to drop the packets at the queue according to packet priority index provided by scalable video coding codec in the application layer. Although the cross-layer QoS approaches mentioned above provide better QoS for video streaming traffic than IEEE legacy DCF and EDCA, they do not provide mechanism to adjust the EDCA parameter sets for different network environment (e.g., the bit-rates of video streaming or the number of uplink and downlink video streaming flows).

4.5. Terminal capabilities

As mentioned earlier, the terminal capabilities can affect the video streaming performance since video decoding itself costs a heavy computing power. Besides, TCP/IP protocols are traditionally implemented in software and executed by the CPU. This approach has numbers of limitations, such as interruptions, memory movement, checksum calculations and fragmentation/reassembly issues. These limitations increase CPU loading while the system processes network applications. When a packet is received, there are about 3100 instructions executed by the CPU and about 50% instructions used for memory copy. If the end-system is an embedded system (e.g., a smart phone), the processing load of TCP/IP protocols would be the performance bottleneck. Thus, a user can not watch the video streaming smoothly and the video quality is bad with fragment frames. Chen et al. [11] design a specific Ethernet network interface card to reduce the overheads of protocol header identification/appending and CRC/checksum calculations. They design a hardware architecture for IP/UDP protocols and checksum. They implement it in the Field Programmable Gate Array (FPGA) prototype of the interface card. As Fig. 13 shows, the hardware architecture is composed of transmission and receiving modules. Video is transmitted by video interface module, DMA controller supports data copy with CPU, transmit data controller monitors buffer queues, UDP/IP encapsulation/decapsulation modules process packet header and UDP/IP recognition module processes payload and does checksum calculation. Using the Ethernet network interface card which operates at 36.6 MHz, the system can speed up video bit stream delivery with a dedicated video interface. Compared with the same operations of a 50MHz ARM processor, the system can save 47,000 ns per frame. Hsiao et al. [12] analyze the operations of TCP/IP protocols and found that the major limitations of CPU operations for video streaming are memory movement, checksum calculations and the large number of interruptions in end-user systems. To resolve the limitations, they propose a dual CPU architecture that accelerates real-time network multimedia transmissions. A FPGA prototype on Versatile board which has an ARM (hard core) and Ini-RISC, is designed and implemented as shown in Fig. 14. The proposed dual CPU architecture outperforms the traditional single CPU system by 37.89% on an FPGA prototype. In order to reduce lost and get higher throughput, Hsiao et al. also propose a high speed multimedia network ASIC to accelerate H.264 streaming [13]. They use slice priority of NAL, partial checksum technique and 750 byte as packet size. The proposed ASIC design can adapt video streaming delivery to limited bandwidth for lower packet losses in high bit error rate wireless environment.

When the H.264 video streaming is received in client side, the factor that affects the CPU loading is the decoding process. The higher resolution of video streaming needs more computing power. Therefore, Guo et al., propose a high-throughput context-adaptive binary arithmetic coding (CABAC) decoder to achieve real-time H.264 streaming decoding [14]. They propose a look-ahead decision

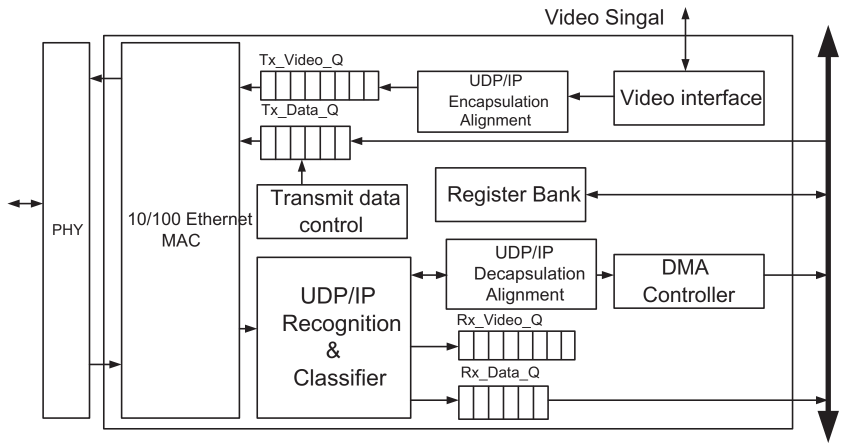

Fig. 13. The designed hardware architecture of IP/UDP protocols in [11].

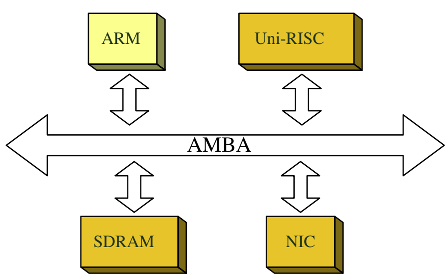

Fig. 14. A dual CPU architecture for H.264 video streaming.

parsing technique on the grouped context table with cache registers, which reduces 62% of cycle count on average as compared with the original CABAC decoding. With high throughput and low-power design, the designed chip is suitable for handheld device. They have integrated the proposed CABAC decoder in a H.264 high profile video decoder system with FPGA verification [15], which passes over hundreds of testing sequences including the conformance sequences from H.264 reference software encoder.

5. Open research issues

Although the works discussed in the previous section try to improve H.264 video streaming delivery in wireless networks from different layer of network protocol stacks, many research issues remain unresolved.

5.1. High performance encoder and decoder forh.264 video coding

H.264 is now the dominant codec for video streaming over wireless networks. Traditional implementations of software for H.264 cannot match real-time streaming requirements. When encoding a Common Intermediate Format (CIF) 30 fps video, the software requires 115.4 MIPS of computing power on a workstation with Ultra Sparc II 1 GHz CPU and 8 GB RAM running SunOS 5.9. However, the average of CPU speed of a smart phone in nowadays is 800 MHz and 1 GB RAM so that is not enough to run the codec in software. The advanced hardware architecture design of the H.264 encoder and decoder can improve the video delivery performance in the application layer, but the design of the architecture still involves a number of challenges. There are many extension profiles of H.264/AVC such as scalable video coding, multi-view video coding and multi-mode power-aware video system. To encode a 3-view 1080p video, 82.4 Theoretical Operations Per Second (TOPS) computing power and 54.6 TB/s memory access are required with a full search algorithm. Thus, view scalability is important for dealing with various prediction structures of 3-D video. Besides, to deliver the multi-view or high definition video over wireless network still needs research efforts since the bandwidth, video smoothing and channel condition are still the limitations in wireless networks.

5.2. Bandwidth estimation and QoS provision in cross-layer design components

Bandwidth estimation is an important component of the cross-layer design architecture because the encoder in the application layer must optimize the encoding behavior based on the available bandwidth for video streaming. Bandwidth estimation is also very useful and important for call admission control (CAC) when the flow of video streaming is set up since the CAC must know the available/residual bandwidth of the wireless networks without affecting the QoS for current flows in the network to check whether the network has enough bandwidth for the new coming video flow. However, existing works focus on bandwidth estimation in IEEE 802.11 DCF. To the best of our knowledge, no works have considered available bandwidth estimation under IEEE 802.11e EDCA, which is the dominant standard for QoS provision in wireless networks. The bandwidth estimation in IEEE 802.11e EDCA wireless networks is really a challenge since it needs to concern both inter-AC and intra-AC interference.

Currently, for QoS provision in 802.11e wireless networks, existing mechanisms map video streaming frames into different ACs or a single AC with different drop probabilities. These mechanisms only adopt the EDCA parameters suggested by 802.11e standard. However, the performance of higher priority AC in the standardized parameter sets is affected seriously by the number of flows of lower priority AC. This calls for QoS provision mechanisms for EDCA parameter configuration for H.264 video streaming in wireless networks to provide guaranteed service for video streaming frames for different network environment (e.g., the bit-rates of video streaming or the number of uplink and downlink video streaming flows), and good performance for other data traffic.

5.3. Resource allocation for multicast services in relay networks

Although Kuo et al have proved that the multicast recipient maximization (MRM) problem with a given resource budget is NP-hard and propose a polynomial-time algorithm to solve it, they

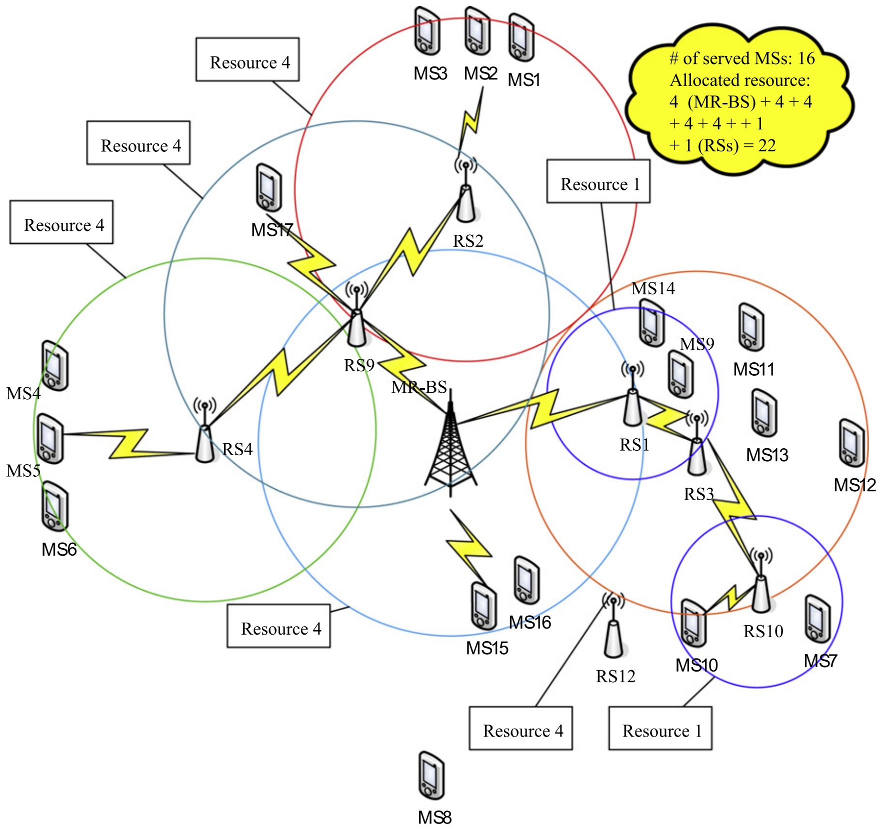

Fig. 15. The optimal solution of Fig. 5 if multi-level relay is considered. 16 MSs are able to be served.

assume that RSs can only relay traffic from SSs. However, in practice, RSs can also relay traffic from other RSs, so the resource allocation can be more efficient. Fig. 15 shows the optimal resource allocation among the MR-BS and RSs if multi-level relay is allowed. We can see that 16 MSs can be served under the same resource budget, which is larger than that (i.e., 14 MSs) in Fig. 8. However, the maximization problem becomes more complex. Besides, the issue of minimizing the resource allocation to serve all SSs for a multicast service in a wireless relay network is not discussed. The MRM problem wants to serve as many MSs as possible within a resource budget. The heuristic strategy is allocating resource to serve MSs with higher utilization first since the budget is limited and not all MSs need to be served. On the contrary, the other problem is to serve all MSs or all members of a multicast group with minimal resource. Since all MSs must be served, the nodes with higher resource to serve may be allocated resource first. Thus utility-based heuristic approaches for MRM problem are not good for the minimization problem. Therefore, these two problems seem similar, but their essence and solutions are quite different. The GA can be applied to these problems since the solutions in GA approaches can be used as the initial populations to calculate the solutions for the following channel conditions. This is because the channel conditions of MSs vary over time, but they should not differ too much in two consecutive time periods. Therefore, unlike heuristic algorithms, GA approaches do not need to execute the whole algorithm to find a new solution, which reduces the computation time.

5.4. TCP/IP protocols offload engine

The computing ability of end-users, especially on mobile devices, is also important in H.264 video streaming. Traditional TCP/IP protocols are implemented in the kernels of operation systems that place a heavy burden on host processors. New hardware architectures, such as multi-core processors, are helpful in end systems, but the IC design cost and power consumption raise other issues. Besides, bit-error resiliency mechanisms can increase throughput that do not drop the bit-error packet. If the resiliency mechanism can be combined with the NAL information of H.264, the system throughput can be improved further.

6. Conclusion

With the rapid growth of the wireless Internet, more and more people are using wireless networks for real-time video applications in end-systems. However, H.264 video streaming over wireless networks presents a number of challenges. Several techniques have been proposed to improve H.264 video streaming transmission in wireless networks, e.g., SVC, cross layer design mechanisms, bit-error resiliency measures, and SoC design for end-user's computing ability. In this work, we classify these approaches based on the network protocol stacks they work on and the SoC design. Finally, we also address some open research issues concerning H.264 video streaming delivery in wireless networks.

References

[1] Kovacs Akos, Godor Istvan, Racz Sandor, Borsos Tamas, Cross-layer quality-based resource reservation for scalable multimedia, Computer Communications 33 (2010) 283–292.
[2] Y.T. Hou, B. Li, S.S. Panwar, H. Tzeng, On network bandwidth allocation policies and feedback control algorithms for packet networks, Computer Networks 33 (2000) 481–501.
[3] Gang Sun, Wei Xing, Dongming Lu, A content-aware packets priority ordering and marking scheme for H.264 video over diffserv network, in: IEEE Asia Pacific Conference on Circuits and Systems, 2008.
[4] Fan Li, Guizhong Liu, A novel marker system for real-time H.264 video delivery over diffserv networks, in: IEEE International Conference on Multimedia and Expo, 2007.

[5] Lishui Chen, Guizhong Liu, A delivery system for scalable video streaming using the scalability extension of H.264/AVC over diffserv networks, in: IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2008.
[6] H. Schwarz, D. Marpe, T. Wiegand, Basic concepts for supporting spatial and SNR scalability in the scalable H.264/MPEG-4-AVC extension, in: IEEE International Conference on Systems, Signals and Image Processing, 2005.
[7] H. Schwarz, D. Marpe, T. Wiegand, Overview of the scalable video coding extension of H.264/AVC, IEEE Transactions on Circuits and Systems for Video Technology 17 (9) (2007) 1103–1120.
[8] W. Liang, Constructing minimum-energy broadcast trees in wireless ad hoc networks, in: The ACM International Symposium on Mobile Ad Hoc Networking and Computing, 2002, pp. 112–122.
[9] D. Li, X. Jia, H. Liu, Energy efficient broadcast routing in static ad hoc wireless networks, IEEE Transactions on Mobile Computing 3 (2) (2004) 144–151.
[10] P.-J. Wan, G. Calinescu, X.-Y. Li, O. Frieder, Minimum-energy broadcast routing in static ad hoc wireless networks, IEEE INFOCOM (2001) 1162–1171.
[11] M.C. Chen, S.F. Hsiao, C.H. Yang, Design and implementation of a video-oriented network-interface-card system, in: IEEE Asia South Pacific Design Automation Conference, 2003.
[12] Yi-Mao Hsiao, Chao-Yuan Wang, Kuo-Chang Huang, Yuan-Sun Chu, An architecture of accelerating real-time multimedia for networking applications, in: IEEE International Symposium on Intelligent Signal Processing and Communication Systems, November, 2007.
[13] Yi-Mao Hsiao, Feng-Pin Chang, Yuan-Sun Chu, High speed multimedia network ASIC design for H.264/AVC, in: IEEE International Conference on Industrial Electronics and Applications, Taichung, June, 2010.
[14] Yao-Chang Yang, Jiun-In Guo, High-throughput H.264/AVC high-profile CABAC decoder for HDTV applications, IEEE Transactions on Circuits and Systems for Video Technology 19 (9) (2009) 1395–1399.
[15] C.-C. Lin, J.-W. Chen, H.-C. Chang, Y.-C. Yang, Y.-H.O. Yang, M.-C. Tsai, J.-I. Guo, J.-S. Wang, A 160K Gates/4.5 KB SRAM H.264 video decoder for HDTV applications, IEEE Journal of Solid-State Circuits 42 (1) (2007) 170–182.
[16] T. Wiegand, G.J. Sullivan, G. Bjøntegaard, A. Luthra, Overview of the H.264/AVC video coding standard, IEEE Transactions on Circuits and Systems for Video Technology 13 (7) (2003) 560–576.
[17] Lars-Åke Larzon, Mikael Degermark, Stephen Pink, UDP lite for real time multimedia applications, in: IEEE International Conference of Communications (ICC), 1999.
[18] Jari Korhonen, Pascal Frossard, Bit-error resilient packetization for streaming H.264/AVC video, in: ACM International Multimedia Conference, 2007.
[19] Chih-Ming Chen, Chia-Wen Lin, Yung-Chang Chen, Cross-layer packet retry limit adaptation for video transport over wireless LANs, IEEE Transactions on Circuits and Systems for Video Technology (2010).
[20] Wen-Tsuen Chen, Tzu-Ching Lin, Yu-Chu Chang, Jyh-Cheng Chen, Dynamic packet selection for H.264 video streaming over IEEE 802.11e WLANs, in: IEEE The Fourth International Conference on Wireless and Mobile Communications, 2008.
[21] Mei-Hsuan Lu, Peter Steenkiste, Tsuhan Chen, Video streaming over 802.11 WLAN with content-aware adaptive retry, in: IEEE International Conference on Multimedia & Expo,2005.
[22] Adlen Ksentini, Mohamed Naimi, Abdelhak Guéroui, Toward an improvement of H.264 video transmission over IEEE 802.11e through a cross-layer architecture, IEEE Communications Magazine 44 (1) (2006) 107–114.
[23] Tao Fang, Lap-Pui Chau, GOP-based channel rate allocation using genetic algorithm for scalable video streaming over error-prone networks, IEEE Transactions on Image Processing 15 (6) (2006) 1323–1330.
[24] J. Thomas Schierl, Cornelius Hellge, Shpend Mirta, Karsten Grüneberg, Thomas Wiegand, Using H.264/AVC-based scalable video coding (SVC) for real time streaming in wireless IP Networks, in: IEEE International Symposium on Circuits and Systems, 2007.
[25] Euy-Doc Jang, Jae-Gon Kim, Truong Cong Thang, Jung Won Kang, Adaptation of scalable video coding to packet loss and its performance analysis, in: IEEE International Conference on Advanced Communication Technology, 2010.
[26] D. Wu, Y.T. Hou, Y.-Q. Zhang, Scalable video coding and transport over broadband wireless networks, Proceedings of the IEEE 89 (1) (2001) 6–20.
[27] Hsing-Lung Chen, Po-Ching Lee, Shu-Hua Hu, Improving scalable video transmission over IEEE 802.11e through a cross-layer architecture, in: IEEE The Fourth International Conference on Wireless and Mobile Communications, 2008.
[28] Chuan Heng Foh, Yu Zhang, Zefeng Ni, Jianfei Cai, King Ngi Ngan, Optimized cross-layer design for scalable video transmission over the IEEE 802.11e networks, IEEE Transactions on Circuits and Systems for Video Technology 17 (12) (2007) 1665–1678.
[29] Yu Zhang, Chuan Heng Foh, Jianfei Cai, An on–off queue control mechanism for scalable video streaming over the IEEE 802.11e WLAN, in: IEEE International Communications Conference, 2008.
[30] T.S. Rappaport, Wireless Communications: Principles and Practices, Prentice-Hall, Englewood Cliffs, NJ, 1996;
G.O. Young, Synthetic structure of industrial plastics (Book style with paper title and editor), in: J. Peters (Ed.), Plastics, second ed., vol. 3, McGraw-Hill, New York, 1964, pp. 15–64.
[31] Wen-Hsing Kuo, Jeng-Farn Lee, Multicast recipient maximization in IEEE 802.16j WiMAX relay networks, IEEE Transactions on Vehicular Technology 1 (2010) 335–343.
[32] Wen-Hsing Kuo, Jeng-Farn Lee, Multicast routing scheme for recipient maximization in wireless relay networks, IEEE Transactions on Vehicular Technology 59 (2010) 4002–4011.