A step-by-step tuning of H.264 for unreliable dynamic networks

C. Calafate, M. P. Malumbres

Resumen— **In this paper we address the problem of achieving reliable video streaming in mobile and unreliable networks such as 802.11b based MANETs. The approach followed was a three step process: general-purpose codec tuning, error-resilience codec tuning and 802.11b MANET evaluation and tuning. In this last step we perform an innovative simulation study based on reproducing the delivery of a live encoded video stream on these networks using the NS-2 simulator. The results focus on packet losses, end-to-end delay, jitter and video distortion. Results show that video traffic has demands that make it difficult to cope with on standard MANETs, and that improvements are required in terms of routing protocols and QoS provisioning either on the MAC layer or at IP level using traffic shaping tools.** [1]

Palabras clave— **H.264, MANETs, 802.11b, video, resilience, routing**

I. Introduction

Increasing use of mobile devices and demand for video oriented applications is leading companies and researchers to look for solutions in the field of mobile multimedia. Several improvements related to video compression technology were made in recent years resulting in the ISO MPEG-4 Part 2[1] standard and ITU-T Recommendation H.263 [2]. During this year the new JVT H.264/MPEG-4 part 10 standard will be finished, offering an enhanced video technology that will provide superior compression performance and better error-resilience, as well as many other features as will be exposed in section III. Such improvements pave the way for ubiquitous human-to-human video communication, even when using low-bandwidth and error-prone network environments.

The widespread deployment of the 802.11b technology, and the advances on other standards of the 802.11 family have increased the interest in MANETs that, although originally intended to cover military or disaster-related situations, are becoming more and more an alternative solution for the enterprise and home environments.

The strategy we followed was to start from a simple configuration where all the options from the codec were disabled by default, then study the impact on performance of each option, and finally choose the option set that offers better performance results. In a second stage the encoder was tuned using the results achieved in the previous one in order to measure the effectiveness of the supported error-resilience options under our simulation framework. In a final stage we use the NS-2[3] simulator to study the performance of a real-time video stream on 802.11b MANETs. We tried to provide an accurate study by focusing on a single H.264 video stream, so that the effects of the different routing protocols and the CSMA/CA radio technology are put into evidence in terms of packet losses, packet loss patterns, end-to-end delay and jitter. At the same time, we will be able to analyze the behavior of the H.264 error-resilience tools, giving us an idea about their effectiveness in terms of perceived video quality distortion.

Concerning the structure of this paper, in the next section we introduce some important aspects related to 802.11b based MANETs. Section III presents the H.264 video codec and the available error-resilience mechanisms, and in section IV we describe the simulation framework. Simulation results are presented in section V, and concluding remarks are made in section VI, along with some guidelines about future work.

II. Issues concerning 802.11b based MANETs

IEEE's 802.11b standard [4] is being increasingly used throughout corporations worldwide due to its good balance of cost, range, bandwidth and flexibility. The bandwidths set by the standard range from 1 to 11 Mbps, but other standards being developed in the same family aim at higher bandwidths, though maintaining the same frequency bands - 2.4Ghz. The 802.11b standard offers operation modes named Point Coordination Function (PCF) and Distributed Coordination Function (DCF). PCF is used in infrastructure mode, where Access Points are responsible for coordinating the transmissions from nodes. DCF, on the other hand, is a distributed mechanism through which each node has the responsibility of sensing the medium, along with avoiding and reacting to collisions.

Our analysis is centered on 802.11b networks with Distributed Coordination Function. When operating in this mode each unicast packet is optionally preceded by a RTS/CTS sequence, followed by a mandatory Acknowledge packet. The RTS/CTS process aims at eliminating the well-known hidden node problem[5]. On the other hand, the acknowledgment assures that the packet, when delivered, is free of bit errors. However, the Link Layer for 802.11b is not connection-oriented, which means that after a standard/user-defined number of failed transmission attempts, a packet is dropped. Also, the link level frames for 802.11b contain error-detection data,

Departamento de Informática de Sistemas y Computadores, Universidad Politécnica de Valencia, Valencia (España). Correo electrónico: `calafate,mperez@disca.upv.es`

[1] Este trabajo se ha realizado en el ámbito de los proyectos CTIDIB/2002/29 y CTIDIB/2002/19 concedidos por la *Oficina de Ciencia y Tecnologia de la Generalitat Valenciana*

which means for all practical purposes that data sent to higher-level layers are free of bit errors. Therefore, the only kind of losses present in such an environment are packet losses.

802.11b's DCF mode enables the formation of mobile ad-hoc wireless networks through node cooperation using one of the available multi-hop ad-hoc routing protocols. Protocols such AODV[6], DSR[7] and TORA[8] are well known ad-hoc routing protocols.

Due to their nature, MANETs are very unstable due to frequent route changes caused by node movement, node on/off activity or even noise.

III. H.264 related issues

The upcoming video coding standard H.264 [9], part of an on-going activity since 1997 named H.26L, is being developed by the Joint Video Team (JVT), an alliance formed by the former ITU-T VCEG and ISO MPEG-4 groups. JVT aims at elaborating an open standard that is not application-specific, and that performs significantly better than the available ISO MPEG-4 Part 2 standard [1] and ITU-T Recommendation H.263 [2] in terms of compression, network adaptation and error robustness.

In H.264 there is a back to basics approach, where a simple design using well known block-coding schemes is used. In the design of this codec, the Video Coding Layer was separated from the Network Adaptation Layer in order to enable a modular development of each of its components. Due to its general purposed nature, some mechanisms were included on both encoder and decoder envisioning enhanced performance in error-prone environments, such as wireless networks or the Internet. By tuning certain parameters, the user can obtain a trade-off between compression rate and error resilience. Such granularity will allow an optimal tuning to the target environment.

Random intra macroblock refreshes and the insertion of intra-coded pictures (I frames) are the most commonly used methods to stop the temporal propagation of errors when no feedback channel is available. While intra frames reset the prediction process, avoiding error propagation, their use has a generally high bandwidth cost, causing also severe bit rate variations. The use of random intra macroblock refreshes is more effective than I frames because it not only aids in generating streams with more constant bit-rate, but can also provide better results by statistically reseting the error for each of the macroblocks.

Multi-frame prediction is another tool targeting to increase both compression performance and error-resilience. This is achieved by using more than one reference frame in the prediction process. As exposed in other work[10], this technique is particularly useful after the loss of a full frame by expecting some of the previous reference frames to be available, enabling partial motion compensation.

Concerning the decoder, it plays a fundamental role in error resilience since it is responsible for error concealment tasks. With that purpose, it keeps a status map for macroblocks which indicates, for each frame being decoded, weather a certain macroblock has been correctly received, lost or already concealed.

IV. Simulation framework

A. Steps on and two

To evaluate the performance of the H.264 codec we used reference software JM3.9a. The evaluation was a three step process, where in a first step the codec was tuned in such a manner that all major options were turned off. The effect of each of the parameters on bit rate, SNR and encoding time were measured in order to provide enough information to tune the codec for step two, where error resilience issues are analyzed.

Due to the absence of a bit rate control mechanism inside the codec, we tuned the codec for mid-range quantization values in order to allow an external bit rate control mechanism that we developed to have the flexibility in range needed for its task. Such mechanism allowed us to perform our tests for step two at a constant bit rate, so that coherence and ease of interpretation was possible.

All tests were run at a frame rate of 10 frames per second, and the test sequences was Foreman in the QCIF format. We chose a mode where only P frames were active at the start of stage two (IPPP...) and since the reference software used did not support errors in B frames, their analysis was not part of our framework.

B. Step three

In order to perform the desired evaluations we used the Network Simulator (NS-2) [3] version 2.1b9a.

NS-2 is a discrete event simulator. The physical layer for the simulation uses two-ray ground reflection as the radio propagation model. The link layer is implemented using IEEE 802.11b Distributed Coordination Function (DCF), and the Media Access Control Protocol (MAC) is CSMA/CA - Carrier Sense Multiple Access with Collision Avoidance. Packets sent by the routing layer are queued at the interface until the MAC layer can transmit them. This queue has a maximum size of 50 data packets, giving higher priority to routing packets. The transmission range for each of the mobile nodes is set to 250m and the channel capacity to 11Mbps (full rate).

The chosen video test sequence is again the well-known QCIF Foreman sequence at 10 frames per second, and the final bit rate for the sequence is 178.64 kbps.

V. Performance Results

In this section we present the most relevant results related to the tuning performed on steps 1 to 3.

Concerning step 3, we start with a square scenario is used in order to take a first measure about the impact of node mobility and network congestion on the final delivered video quality. These are two different aspects that may affect the video quality performance at different degrees. For that reason we proceed by testing node mobility and network congestion independently. Finally, we will test the behavior of the H.264 video codec at different network congestion levels in order to analyze the effectiveness of its error-resilience tools on MANETs.

A. Step 1: H.264 issues

Concerning the main parameters of interest after step one, the Hadamard transform, CABAC and Rate Distortion Optimization were turned ON since they presented the best results. The use of adaptive block transforms for inter and intra was set to the fully flexible mode taking into account the results available in [11].

JM3.9a reference software capabilities concerning RTP packetization and NAL encapsulation were activated in order to simplify the tasks relative to packet loss evaluation. The packetization process was set to 7 packets per frame after an initial evaluation presented in [12] and taking into account the characteristics of ad-hoc networks.

B. Step 2: Evaluation under random packet losses

On this subsection we will present the performance results of the parameters considered most relevant in random error scenarios. The purpose of these results is to allow the user to tune the codec according to the expected channel errors. It should be pointed out that all values are average values of ten consecutive simulations.

Figure 1 presents the achieved results by varying the intra frame period. As it can be seen, smaller intra frame periods perform better in scenarios with higher packet loss. These results can be directly compared to those presented in figure 2 relative to random intra macroblock updates since the conclusions are similar. In terms of quality, doing Intra Macroblock Updating improves the codec performance in medium to high motion sequences for low and moderate packet loss rates (up to 3%). However, this improvement is based on low updating rates, avoiding to increase the final bitrate significantly. As exposed in [13], it is interesting to dynamically change the macroblock update frequency depending on the expected packet loss rate in order to maximize both quality performance and error-resilience features.

The results for the Macroblock Lines Intra Update parameter and the use of SP frames mimics the results of figures 1 and 2, therefore they were not presented here. For more information refer to [12]. In general, all of these four parameters have similar effects regarding error resilience, and tuning one of them, as the Random Intra Macroblock Updates parameter, provides satisfactory results.

Concerning the results relative to Constrained Intra Prediction, these show that its use outperforms the original sequence when error surpasses the 4-5% to an excess than can reach 0,7 dB.

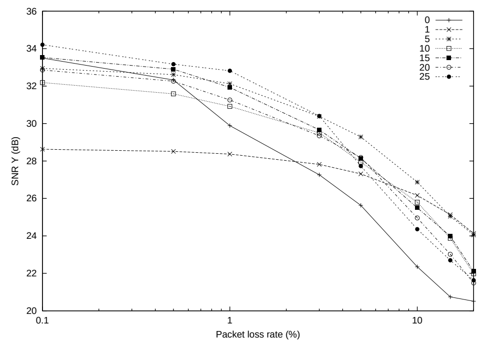

Fig. 1. PSNR results for different I frame intervals.

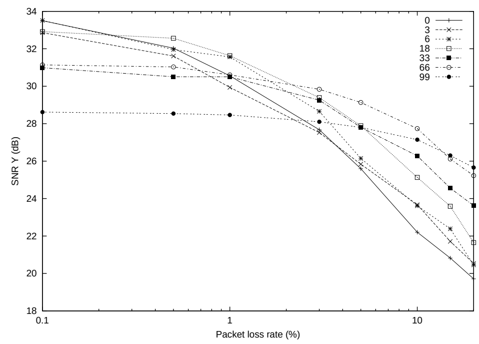

Fig. 2. PSNR results for varying Random Intra Macroblock Updates.

C. Step 3: 802.11b MANET evaluation

C.1 Mobility evaluation

We devised a 30 node scenario in a 670x670 area. Mobility was generated through the random waypoint model available in the NS tool and tuned to a speed of 2 meters per second, followed by a wait time of 5 seconds for all nodes. In addition to the video flow, 5 background FTP flows are also set (1 every 6 nodes). Figure 3 shows the results achieved by using different routing protocols, with this scenario, in terms of distortion and packet loss rate.

“Hello” based AODV (AODVUU-H) performs well in situations of very low mobility because route changes do not occur so often. Also, there are less chances that background congestion causes one link to be considered lost (3 consecutive “Hellos” have to be lost). Its link aware counterpart (AODVUU) performs well on low and average mobility scenarios. TORA shows the best overall behavior under this scenario, showing good distortion levels at all speeds and good ability to maintain the packet loss rate at high speed. DSR is also able to maintain steady levels of distortion and packet loss rate, although not so efficiently as TORA.

This analysis does not pretend to be an in-depth examination of these routing protocols, but rather a study on best choices to achieve good video robustness. Please refer to works such as [14] for a more general study on the performance of different routing protocols.

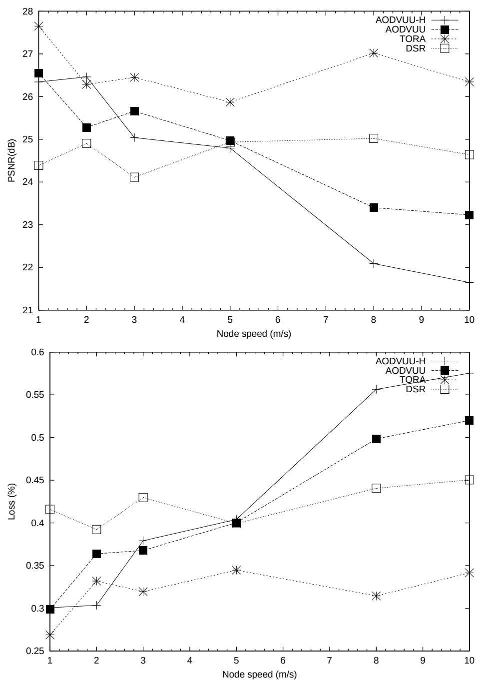

Fig. 3. Evaluation of different routing protocols for varying mobility in terms of packet losses and perceived PSNR for the video test sequence

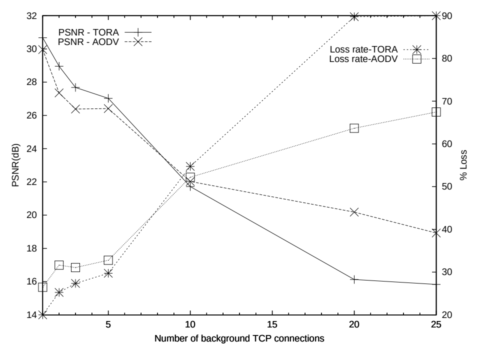

Fig. 4. PSNR and packet loss rate performance for a variable number of background TCP connections

C.2 Performance under congestion

After the mobility evaluation we chose both TORA and the AODV implementation from Uppsala to proceed with our analysis. We evaluate their performance when submitted to different levels of congestion at user mobility levels (the value used was 2 m/s for speed and 5 seconds for wait time as previously).

These results were achieved using the same 30 node square scenario described in the previous subsection.

Figure 4 allows us to compare the performance of TORA and AODV with a variable number of TCP connections in the background (TCP traffic is currently the most common - FTP, Web, Telnet, Database Access, etc. - but perhaps not in a near future). We can see from that figure that acceptable distortion levels cannot be reached with more than 10 background connections using either TORA or AODV. TORA is, therefore, the best choice for this range and, even though AODV performs significantly better under critical levels of congestion, the results in terms of distortion are almost at noise levels.

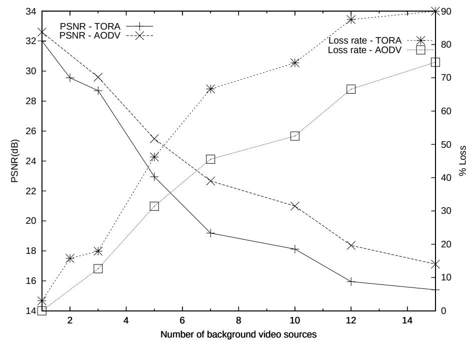

Fig. 5. PSNR and packet loss rate performance for a variable number of background video connections

Figure 5 performs a similar analysis, but now all the background traffic is composed of video flows identical to the one under evaluation. In this scenario AODV always performs better than TORA and, in overall, we consider AODV to be an adequate choice to support video flows as reliably and uninterruptedly as possible.

C.3 Results on the effects of re-routing and background traffic

To complete our analysis, we altered the scenario so that it maintained the same number of nodes and area as before, but was made rectangular (1500x300) to increase the average number of hops. Envisaging a differentiated analysis of mobility and congestion, we started with a situation having neither background traffic nor mobility. We then analyzed separately the effect of allowing high mobility to all nodes (10 m/s and no background traffic) and the effect of congesting the network by setting all the nodes to transmit a moderated amount of CBR traffic (no movement). In all situations, the average (or exact) number of hops was three; the routing protocol used was AODVUU.

Figure 6 shows the effects of mobility and congestion on user perceived distortion. As can be seen, mobility affects distortion in a bursty fashion, typically causing the loss of multiple frames and consequently freezing the image. On the other hand, traffic congestion causes packets to be lost in a more random fashion, so that distortion variation is smoother though more frequent.

The delay analysis also evidences the nature of both kinds of losses, as presented in figure 7.

In the reference situation, more than 99,9% percent of the packets arrive before 7 ms; with high mobility, 92% of the packets arrive in less than 10 ms. Point X is the frontier of two distinct regions: the one on the right where a very small number of packets have very high delays (as much as 6 seconds or more), and the one on the left where packet forwarding is uninterrupted. In the “mobility” scenario, although the average number of hops is 3, this value varies throughout the simulation, explaining why some of

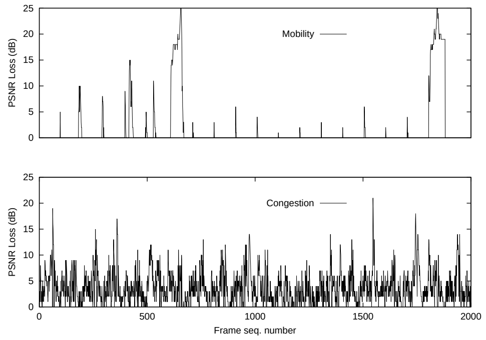

Fig. 6. Effect of congestion and mobility on user perceived PSNR

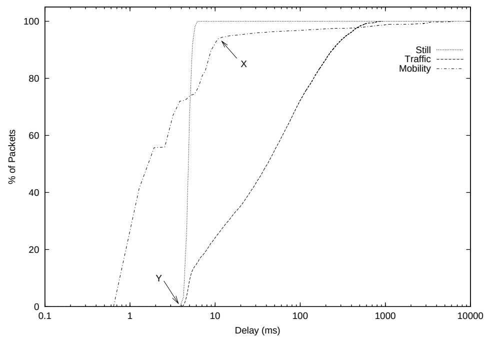

Fig. 7. Delay effects of congestion and mobility

the packets arrive earlier than those in the reference scenario and others arrive later (before X). The phenomena whereby some packets arrive with very high delays (after X) is expected since AODV causes packets to wait in a queue when re-routing tasks are being performed.

Congestion causes a very different behavior, so that all packets that arrive at the destiny do so in less than 1 second, though the delay between consecutive packets can vary greatly. The start point (Y) for both reference and congestion scenarios is common because the destination is 3 hops away on both.

The jitter analysis of figure 8 also aids at visualizing the behavioral difference between both. Even though the jitter peaks occur rather infrequently, they are an order of magnitude superior than those caused by congestion. We conclude that jitter peaks usually translate into a change of route when using reactive protocols.

As it could be inferred from previous results, tightening the limits on packet delay causes more packets to be dropped in high-congestion scenarios than in high-mobility ones. However, these effects can be countered by QoS policies at either the MAC or higher levels. Transmission breaks due to mobility are much more difficult to counter and are more critical. Solutions to this problem could be introduced at the MAC level itself, either by giving routing traffic a higher priority or by assuring that collisions are minimal. The later can be achieved by assigning a dedicated MAC inter-frame space to it, requiring new changes to the specification. Neither of these options, though, are able to provide a delivery guarantee even to a single surrounding node.

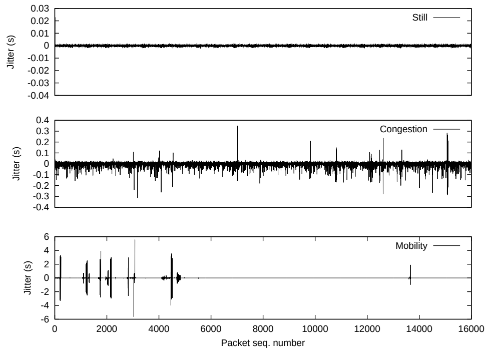

Fig. 8. Jitter due to congestion and mobility

C.4 Evaluation of video codec choices

Our evaluation concerning the video codec parameters focuses on two topics: the number of reference frames and the best method for intra-macroblock updating.

Concerning intra-updating of macroblocks, H.264 provides several choices to the user. We have evaluated the main available choices in the reference software, which are: use of I frames, intra update a predefined number of macroblocks randomly and intra update a whole line randomly chosen for each frame.

The scenario is the same one used in previous subsections. It is considered as an example of high congestion, with a packet loss ratio of 20%.

Table I presents the average distortion values for this scenario. The use of random macroblock updates proves to be the best option in terms of error-resilience, showing its effectiveness with respect to no updating (almost 5 dB of difference).

TABLA I

AVERAGE PSNR RESULTS EVALUATING STRATEGIES FOR INTRA MB UPDATING

Updating method	Avg. PSNR (20% loss)
1/3 random MB updates	25,58
IPP sequence	24,01
IPPPPP sequence	23,35
Random line intra update	22,79
No intra MB updates	20,62

The process of random intra-macroblock updating could be tuned to adapt to network congestion interactively. This process would require a feedback channel, which would also increase network congestion. Therefore, we didn't consider it a priority, though certainly a possibility.

Evaluation relative to the number of reference frames was done in the situation it was originally proposed for: entire frame losses. Instead of running a high mobility scenario (known to cause that kind of losses), we have directly tested the effects of loosing 1 to 5 consecutive frames, so that the error propagation effect was presented as clearly as possible. Results shown in figure 9 evidence the appreciated

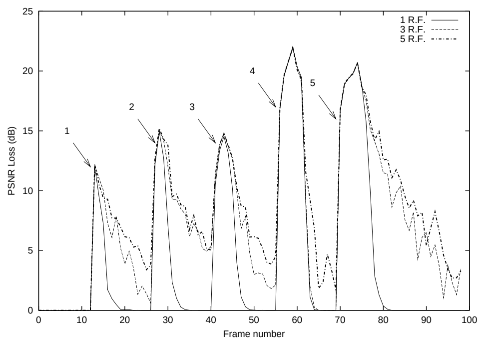

Fig. 9. Analysis of error propagation by simulating an increasing number of entirely lost frames

behavior when using 1, 3, and 5 reference frames. Number/arrow pairs refer to how many frames were lost. From figure 9 we can conclude that using a single reference frame is the most effective choice to stop temporal error propagation. A proper tuning choice for H.264 in MANETs would be, therefore, to use only one reference frame; demands in terms of memory on both encoder and decoder are also reduced by this setup.

We are not aware if the result we obtained is due to an error in the implementation of H.264 codec, though we believe not.

VI. Conclusions and future work

We presented the main issues related to 802.11b based MANETs, taking into account the requirements of real-time video.

The random-loss results obtained allow tuning the encoder according to the expected packet loss rates inside the network, and show that a careful choice can increase significantly the overall PSNR of the sequence. Assuming a typical situation where there is a 10% packet loss in the network, tuning the Random Intra Macroblock Update to 1/3 of the frame size improves error-resilience on random and burst error situations at the cost of only a marginal increment in bitrate.

We proceeded by simulating real-time video on MANETs under average congestion, where TORA has shown to offer the best distortion results to the video stream. Variable congestion tests followed using TORA and AODVUU. Using TCP as background traffic, TORA has only provided slightly better results with less than 10 connections, with AODVUU offering a better overall performance. In fact, up to four extra video connections can be achieved with AODVUU relative to TORA maintaining the same level of distortion.

Scrutiny of our results evidenced that even though routing protocols detect broken links in milliseconds, they are not able to perform re-routing tasks as quickly as would be desired. This phenomena occurs because, due to collisions, they are not always able to successfully broadcast routing packets, causing long transmission breaks. In fact, increasing background traffic intensifies this problem, causing routing tasks to become more and more infeasible.

An analysis of delay and jitter followed, showing the effects of congestion and mobility on video streams separately. Here, the ON/OFF behavior with high mobility has shown to cause the loss of communication during long time periods (i.e., 10 seconds or more), being therefore prone to cause annoyance to the receptor. This point will require special consideration in further enhancements.

Concerning the H.264 video codec, we have also showed that the tuning performed on step 2 was effectively resilient in terms of macroblock updating. The use of more than one reference frame, though effective in reducing bit-rate, increases the temporal error propagation and it should be avoided, expect for situations where the media is reliable (CD, DVD, or hard-disk).

Future work will focus on finding techniques suitable for offering good QoS to video streams by differentiating traffic flows, as well as by using multi-path techniques.

Referencias

[1] "ISO/IEC IS,Coding of Audio-Visual Objects, part 2: Visual (MPEG-4)," Information Technology, November 2001.

[2] "Video coding for low bitrate communication," ITU-T Recommendation H.263, 1995.

[3] K. Fall and K. Varadhan, "NS notes and documentation," The VINT project, UC Berkeley, LBL USC/ISI, and Xerox PARC, November 1997.

[4] IEEE/IEC Std 802.11, *Wireless LAN Medium Access Control(MAC) and Physical Layer (PHY) specifications*, The Institute of Electrical and Electronics Engineers, Inc., August 1999.

[5] F. A. Tobagi and L. Kleinrock, "Packet switching in radio channels: Part-II - the hidden terminal problem in carrier sense multiple-access models and the busy-tone solution," *IEEE Transactions in Communications*, , no. COM-23(12):1417143, 1975.

[6] Charles E. Perkins and Elizabeth M. Royer, "Ad hoc On-Demand Distance Vector Routing," in *Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications, New Orleans, LA*, February 1999, pp. 90–100.

[7] David B Johnson and David A Maltz, "Dynamic source routing in ad hoc wireless networks," in *Mobile Computing*, Imielinski and Korth, Eds., vol. 353. Kluwer Academic Publishers, 1996.

[8] V. Park and S. Corson, "Temporally-ordered routing algorithm (TORA) version 1 - functional specification," Internet Draft, MANET Working Group, draft-ietf-manet-tora-spec-03.txt, November 2000, Work in progress.

[9] "Joint Committee Draft J-CD," http://bs.hhi.de/ wiegand/JointCommitteeDraft.pdf, May 2002.

[10] M. Budagavi and J. D. Gibson, "Error Propagation in Motion Compensated Video over Wireless Channels," *Proceedings of the IEEE International Conference on Image Processing, Santa Barbara, USA*, pp. 89–92, October 1997.

[11] Till Halbach and Mathias Wien, "Concepts and performance of next-generation video compression standardization," *5th Nordic Signal Processing Symposium (NORSIG-2002)*, October 2002.

[12] Carlos Miguel Tavares Calafate, "Evaluation of the H.264 encoder (internal report)," DISCA, UPV, Spain, 2003.

[13] Thomas Stockhammer, Dimitrios Kontopodis, and Thomas Wiegand, "Rate-Distortion Optimization for H.26L Video Coding in Packet Loss Environment," *12th International Packet Video Workshop (PV 2002), Pittsburg, PY*, May 2002.

[14] S.-J. Lee, C.-K. Toh, and M. Gerla, "Performance Evaluation of Table-Driven and On-Demand Ad Hoc Routing Protocols," in *Proceedings of IEEE PIMRC'99, Osaka, Japan*, September 1999, pp. 297–301.