

Energy and Distortion Analysis of Video Compression Schemes for Wireless Video Sensor Networks

Saeed Ullah¹, Junaid Jameel Ahmad², Junaid Khalid¹, and Syed Ali Khayam¹

¹School of Electrical Engineering and Computer Science (SEECS),

National University of Sciences and Technology (NUST), Islamabad, Pakistan

²Department of Computer and Information Science, University of Konstanz, Germany

Email: {saeed.ullah, junaid.khalid, ali.khayam}@seecs.edu.pk, junaid.ahmad@uni-konstanz.de

Abstract—Wireless Video Sensor Networks (WVSNs) - a type of WSNs - comprise of sensor nodes that can capture, process and communicate video frames. The battery powered sensor nodes have limited hardware resources while video processing and communication are resource intensive tasks i.e., require high-end processors, large memory and bandwidth. Video encoding is a popular method used to reduce the communication overhead but being an inherently complex process it results in higher computational energy-drain on video sensor nodes. This establishes an interesting computation-communication tradeoff for energy efficient video communication (encoding and transmission) in WVSNs. In this paper, we study this computation-communication tradeoff under Intel-imote2 based single-hop and multi-hop video sensor networks tested by empirically evaluating selected implementations of the MPEG-4 (Part 2) and H.264/AVC encoders. The analysis has been carried out to characterize the performance of encoders in terms of energy efficiency, compression efficiency and video distortion. The experimental results show that in single-hop WVSNs, MPEG-4 is energy efficient over H.264 whilst in multi-hop WVSNs, H.264 is energy efficient over MPEG-4.

Index Terms—Video compression/communication, Wireless Video Sensor Networks (WVSNs), Energy efficiency.

I. INTRODUCTION

With rapid advancement in Micro-Electro-Mechanical Systems (MEMS) technology, the dream of having low-cost sensor nodes is quickly becoming a reality. The sensor nodes can sense, process, communicate wirelessly and even react to their environment. Consequently, when deployed in a region of interest, they are capable of forming a communication network. It is important to note that sensor nodes are battery powered devices and are often deployed in the environments where charging or changing the battery is not always feasible. Furthermore, wireless sensors often carry small memory, a low-end processor and low-bandwidth transceivers. Therefore, efficient utilization of battery and hardware resources have always been the most important aspects considered in development of WSN's applications and protocols.

Wireless Video Sensor Networks (WVSNs) [1] is a category of WSNs in which sensor nodes are equipped with a digital camera. Therefore, they are capable of capturing, processing and communicating multimedia contents in real time. However, in contrast to the limited hardware resources available in

WSN nodes, video processing applications require more powerful hardware (processing and memory) and communication (bandwidth) resources. Due to these limiting factors, it is an extremely challenging task to realize practical WVSNs.

For video communication in WVSNs, video data is often compressed before transmission. Hence, total energy drain for video communication will be the sum of encoding computational energy (i.e., energy depleted for local computations during video encoding) and compressed bits transmission energy. This models the total energy for video communication to be a function of computation-communication tradeoff. Furthermore, the practical requirements of having higher quality levels for the relayed video and multi-hops between the source and destination in WVSNs introduce the challenges of larger bitstreams and data retransmissions, respectively. Coping these two challenges can increase the total energy consumption even further.

Several efforts have been made recently to realize the dream of practical WVSNs. In [2], [3] and [4], the authors have proposed modification in video/image compression algorithms to make them portable for energy constrained devices. In [5], authors evaluated various video encoders on stargate based WVSNs testbed. However, this work only considered single-hop communication scenario and also did not evaluate varying video quality levels. Therefore, to better quantify which available encoder is suited for a particular application in a particular environment, available video encoders should be investigated for this computation-communication tradeoff under varying video quality levels and number of network hops. The main objective of this paper is to perform an intended, unbiased and extensive empirical analysis of existing video encoders for the computation-communication tradeoff along with the associated parameters of video quality and number of hops of realistic WVSNs testbed.

Two video encoding techniques the MPEG-4 (Part2) [6], [7] and H.264/MPEG-4 AVC [8], [9] are used for experiments reported in this paper. Our experimental results show that under single hop communication scenario, MPEG-4 coding is generally more energy efficient than H.264 and intra coding is the most energy efficient configuration. On the other hand,

for multi-hop communication scenario, H.264 being a better compressor than MPEG-4 and inter coding configuration is more energy efficient configuration than intra coding. Furthermore, for the videos encoded at higher quality, H.264 exhibits to be more energy efficient than MPEG-4.

The rest of this paper is organized as follows. Section II covers communication aspects in WVSNs. Evaluation framework is presented in section III. Section IV provides a detailed discussion of the experimental results. Conclusion and future work is discussed in section V.

II. COMMUNICATION MECHANISM IN WVSNS

WSNs can be configured to operate in two types of communication scenarios; single-hop and multi-hop. In single-hop communication scenario, the base station is at one hop transmission distance from the sensor nodes. In a multi-hop communication scenario, sensor nodes and the base station are separated by multiple hops. In this configuration, intermediate sensors act as routers, which forward packets towards the base station. Therefore, in WVSNs deployments (either as single-hop or multi-hop network) there has to be one time video capturing, encoding and decoding. The capturing and encoding processes are done at the source sensor node and the decoding process is performed at the base station, which normally is a resource-sufficient device. Furthermore, if N is the number of hops between the source node and the base station, the encoded bitstream is transmitted N times and received $N - 1$ times before it is received at the base station.

A. Communication Overhead

In WVSNs, whenever a node captures multimedia contents it sends these contents to the base station where a global assessment is performed. CSMA/CA protocol is followed to fairly access shared wireless medium. In multi-hop communication scenario, multiple sensor nodes share the channel which may lead to collisions at the receiving end. To minimize the number of collisions, CSMA/CA algorithm uses the concept of backoff periods – one backoff period equals the time required for transmitting 20 symbols where each symbol is equal to the time required for 4 bits transmission. The CSMA/CA algorithm is shown in Fig. 1. According to [10], MAC packet transmission begins with a CSMA wait for a random number of backoff periods ranging from 0 to $2^{BE} - 1$, where Backoff Exponent (BE) has value between $macMinBE$ and $macMaxBE$ (with default values of 3 and 5, respectively). After the expiry of backoff period, the node performs Clear Channel Assessment (CCA) over a time duration of 8 symbols to determine whether the channel is idle or busy. If the channel is busy, node increments BE by one (if it is less than $macMaxBE$) and enters the backoff period again. If the CCA fails for M number of times, which equals to 4 by default, then packet is discarded. When the CCA succeeds, the node performs an RX-to-TX turnaround for a time period of 12 symbols and transmits the packet. By receiving the packet successfully, the receiving node sends an 11 bits acknowledgment (ACK) back to the sending node. If the sending node does not receive

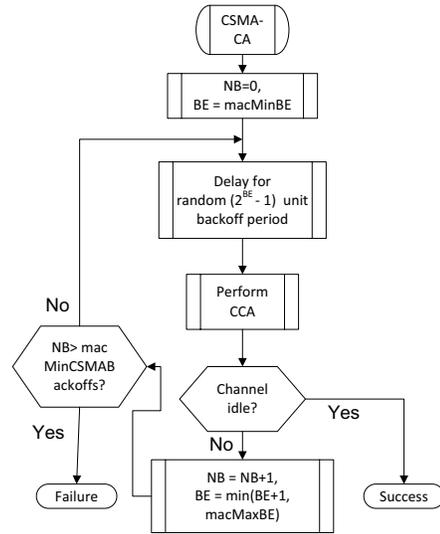


Fig. 1. Un-slotted CSMA/CA Algorithm.

any ACK in $macAckWaitDuration$ (54 symbols), it assumes that the packet is lost and re-sends the packet. Although the CSMA/CA quite heavily reduces the probability of collisions but it cannot completely avoid collisions.

Our sensor network testbed is comprised of Intel-mote2 (imote2) motes. Imote2 have CC2420 radio transceiver, which supports a maximum data rate of 250 kbps. According to the CSMA/CA algorithm, on the 250 kbps network total wait time required between sending two successive data packets is 2.60608 milliseconds.

B. Energy Wasted in Collision and Re-transmission

In CSMA/CA, a node transmits a packet only if it senses an idle channel. Collisions are still possible due to hidden node problem.

The collision probability due to the existence of a hidden node is computed in [11] as given below.

$$P_{coll(h)} = 1 - \frac{(e^{\rho_A} - 1) - \frac{k\rho_A\rho_C}{\rho_A + k\rho_C}}{(e^{\rho_A} - 1) \left(e^{\rho_C} + \frac{\rho_C}{\rho_A} \right) - \frac{k\rho_C}{\rho_A + k\rho_C}} \quad (1)$$

where ρ_A and ρ_C are transmitting and hidden node's traffic loads, respectively, which are equal in our case. Constant k is defined as

$$k = 1 + \frac{W_0(-\rho_C e^{-\rho_A - \rho_C})}{\rho_C} \quad (2)$$

where $W_0(z)$ denotes the Lambert function, which is the solution of the equation $xe^x = z$. For further details, interested readers are referred to [11].

III. EMPIRICAL EVALUATION FRAMEWORK

In this section, we present details of our empirical evaluation framework, which is used to conduct the experiments.

A. WVSN Testbed

The topology of our WVSNs testbed is shown in Fig. 2. The maximum possible distance of any node from the base station can be up to 5 hops (i.e., $N = 5$). Our WVSNs testbed

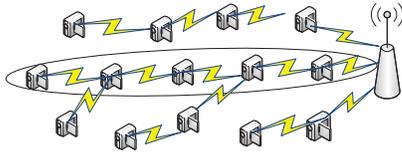


Fig. 2. Imote2 WWSN Testbed.

consists of Intel-mote2 [12] devices. Imote2 is an advanced wireless sensor node platform by Crossbow. It has low power PXA271 XScale CPU and integrates a CC2420 IEEE 802.15.4 radio transceiver. The PXA271 is a multi-chip module that includes three chips in a single package – the CPU with 256kB SRAM, 32MB SDRAM and 32MB of FLASH memory. It also includes a wireless MMX coprocessor to accelerate multimedia operations. The integrated radio transceiver supports a data rate of 250kbps with 16 channels in the 2.4GHz band and the transmission range is about 30 meters. For our experiments, we installed embedded Linux OS on imote2 and ran all experiments on the Linux shell. We used the arm-linux-gcc-3.4.1 compiler to build the C/C++ implementations of the evaluated codecs.

B. Energy Measurements

For all the video communication scenarios evaluated in this paper, the total energy consumed to communicate the video from the source node to the base station is the sum of the energies depleted for encoding computations, bitstream transmissions and receptions¹. The computational energy E_{comp} for video encoding is calculated using Equation 3.

$$E_{comp} = V * I * T_{enc}. \quad (3)$$

Voltage V is measured using voltmeter, which was 4.95V for all the experiments, current I represents the current flow through Imote2 during the video encoding process and is measured using a PC based Oscilloscope [13], and the T_{enc} is the time taken by the motes to encode the video. Note that, I varies based on the content of the video data being encoded. The Oscilloscope stored the current measurements in the form of a graph. An example of one such graph is shown in Fig. 3. We processed current measurement graphs in Matlab to calculate the average current drawn by the mote during video encoding process.

Energy consumed for communication (transmission or reception) of the compressed bit is calculated using Equation 4.

$$E_{comm} = V * I * T_{(T_x/R_x)} \quad (4)$$

In our experiments, voltage V drawn by the Imote2 for transmission was the same as it was in Equation 3. To transmit at 0dbm, CC2420 draws a current of 17.4mA. Furthermore, to operate at 104MHz, Imote2 draws a current of 66mA [12]. The sum of these two currents results in the total transmission current. Similarly, CC2420 draws 18.4mA current in the reception (R_x) mode. Finally, $T_{(T_x/R_x)}$ is the time taken by

¹In this work, we did not consider the energy depletion for video capturing. All videos were already captured and available in main memory of the motes before we started the benchmarking of video encoders.

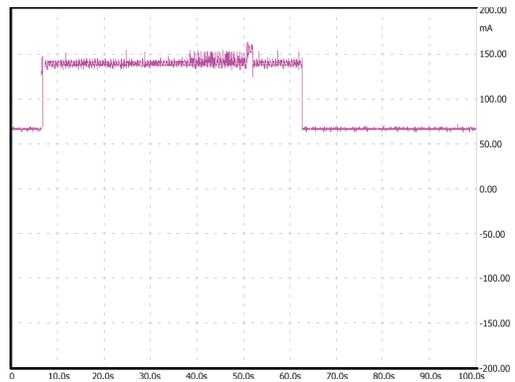


Fig. 3. Current drawn by sensor during encoding

the mote for transmitting/receiving the encoded bitstream and was calculated by dividing the size of the encoded bitstream with the data rate (250kbps). Since the encoded bitstream was sent as 802.15.4 MAC packets, we also considered 9 bytes header for every 93 bytes payload [10]. Equation 4 provides the transmission energy over a single hop. For multi-hop communication, the total transmission energy for N hops E_{commN} is given by Equation 5.

$$E_{commN} = E_{T_x} * N + E_{R_x} * (N - 1) \quad (5)$$

where E_{T_x} and E_{R_x} are energies consumed by a node for transmission and reception of compressed bitstream, respectively.

In this paper, we assume that each node have 4 nodes in its transmission range and generates a traffic of 8 packets per second. By solving Equations 1 and 2 with the above assumptions and adding the packet loss due to physical layer corruption, we are resulted with a total of 30 percent packet retransmission.

C. Selected Video Codecs and Sequences

In this paper, we evaluated two prominent video encoders – MPEG-4 (Part 2) [6] and H.264/AVC [7], [8]. For MPEG-4, we used XVID's implementation [14] and for H.264, we used VideoLAN's implementation namely x264 [15]. There are two variants of predictive video encoding – inter-coding and intra-coding. Inter-coding exploits the redundancy among different video frames using the motion compensation/prediction. Intra-coding, on the other hand, forgoes motion compensation/prediction block and encodes each frame as an image. The effects of motion compensation/prediction will be elaborated in the subsequent section.

We used different types of video sequences to cover diversity of video content so that video communication over WWSN does not remain biased for any particular type of video sequence. For our experiments, we used 3 standard video sequences: Carphone, Foreman and Mobile. All of these sequences were in the YUV format of QCIF resolution (176×144). These videos have different motion characteristics. In carphone sequence the camera is stationary, there is very little motion in the background and relatively more motion in the foreground. Foreman sequence has some motion

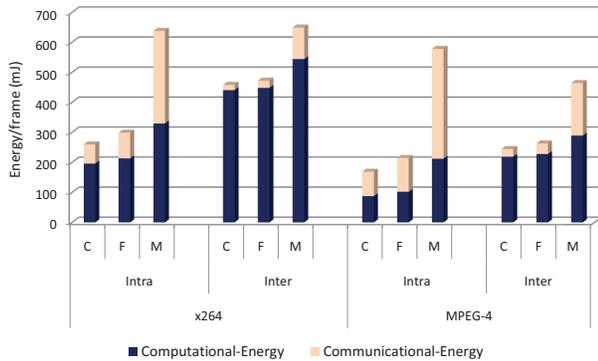


Fig. 4. Inter and intra frame encoding on single-hop at 35db PSNR

in the background, sharp motion in the foreground and a little motion in the camera. Mobile sequence is the most complex of the other two as it has continuous object and the camera motion.

For all experiments, we encoded first 96 frames of each video sequence at 12 *fps*. All the energy measurements are reported on per frame basis, unless otherwise stated. For the inter encoding, we used GOP size of 6 with frames type sequence $I-P-B-P-B-P-I$. Encoding was performed using the main profile for x264 and simple profile for MPEG-4. Furthermore, distortions are reported as Y-PSNR only. In order to make the results more comprehensive, we presented results at two quality levels 30dB and 40dB PSNR.

IV. EXPERIMENTAL RESULTS

This section presents the results of our study. In figures, we acronymed Carphone as C, Foreman as F, Mobile as M. First part of this section presents evaluation of encoders in single-hop communication scenario of WVSNs whilst second part presents the evaluation for multi-hop communication scenario.

A. Single-Hop WVSNs

Figure 4 evaluates the inter and intra coding. If we look at the total energy consumption break down, we see that inter coding consumed more energy for encoding computation than intra coding. This is because of the presence of motion-compensation/prediction block in inter-coding. This is the reason why, inter-coding generally achieved better compression ratio and thus consumed less energy in communicating bitstreams. In contrast, intra-coding saved considerable computational energy (because it does not perform the motion-compensation/prediction). However, intra-coding did not result in comparatively better compression ratio and had to pay more in terms of communication energy. In single-hop communication scenario, for both (x264 and MPEG-4) intra-coding generally remained efficient over inter-coding. However, for video sequences like Mobile, which have continuous motion throughout the scene, the computational energy used for motion-compensation/prediction did not help in achieving better compression ratio. This resulted in higher energy drain for both computations and bitstream transmission. In addition, regardless of the video encoder, both computational and

communication energy drains were observed to be dependent on the complexity of the video content and affected the overall energy required to communicate videos.

It should also be noted that both inter and intra coding configurations of MPEG-4 are energy efficient over x264. Despite the fact that x264 achieved better compression ratio over MPEG-4 (see Fig. 4), when it comes to the total energy depletion, the computational complexity of x264 made it inferior to MPEG-4. Based on these evaluation results, we can suggest that MPEG-4 based intra coding is efficient configuration for single-hop video communication.

B. Multi-Hop WVSNs

In single-hop communication, we established that x264 achieved a better compression ratio but it was computationally expensive over MPEG-4. Furthermore, inter-coding always resulted in better compression ratio over intra-coding. In multi-hop communication scenario, video encoding is performed once and bitstream has to travel a number of hops to reach base station. A video codec configuration, which have better compression ratio should be a better choice for video communication in multi-hop WVSNs. In addition, the decision of changing the bitrate can seriously effect (increase/decrease) the amount of communication energy in a multi-hop network. Therefore, in order to better evaluate the behavior of video communication in multi-hop WVSNs, we selected a multi-hop network with 5 nodes and performed our performance evaluation at two fixed video quality levels: 30dB and 40dB of PSNR.

1) *Inter vs. intra coding*: Figure 5 plots the accumulated energy depletion at each node of our multi-hop video sensor networks. The focus of plotting the energy curves in Fig. 5 was to separately compare the intra and inter encoding of x264 and MPEG-4 encoders at two video quality levels. By looking at the plots in Figures 5(b)-5(d), we observed that regardless of the video quality level, intra-coding was initially better over inter-coding for relatively less general purpose video sequences (Foreman and Carphone). From 2nd hop of the network, the higher compression ratio of inter coding made it to become energy efficient over intra coding. However, the same two video sequences when encoded for 30dB via x264 (see Fig. 5(a)), intra coding remained better until 5th node of the network. It should be noted that the gap between the inter and intra energy curves decreases as the network size increases. Based on this trend, we can assume that inter-coding would possibly become energy efficient over intra-coding at 6th or 7th node. On the other hand, for complex video sequences like Mobile, again inter coding failed to achieve higher compression ratio. Based on the evaluation results of Figures 5(a)-5(d), we can in general suggest to use inter coding in multi-hop networks.

2) *Cross comparison of encoders*: In Fig. 6, we re-plot the energy curves of Fig. 5 to present an comparison between x264 and MPEG-4 video codecs. Fig. 6(a) and Fig. 6(b) present the evaluation of the inter-coding of x264 and MPEG encoders at two quality levels. For relatively less complex

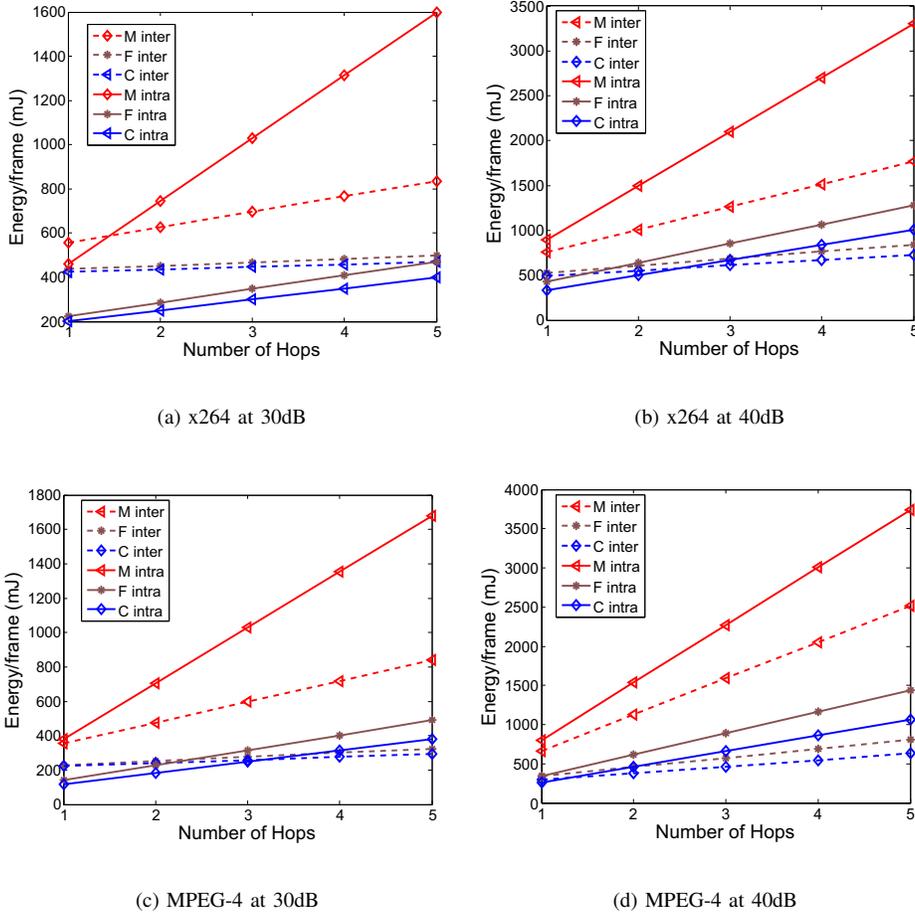


Fig. 5. Comparison of inter and intra encoding

video sequences, it should be noted that MPEG-4 is more energy efficient than x264. This is because the bitstream size resulted by inter coding for these video sequence was generally small and comparatively higher encoding computations for x264 still dominate its total energy drain. However, for a complex video sequence like mobile (for which inter coding failed to produce smaller bitstream), the energy curves dictates MPEG-4 based inter coding to be more energy efficient over its corresponding x264 based inter coding.

Similarly, Fig. 6(c) and Fig. 6(d) present the energy curves for the two quality levels of the x264 and MPEG-4 based intra codings. Since x264 based intra-coding has slightly better compression ratio over MPEG-4 based intra coding (see Fig. 4), in contrast to the single-hop communication scenario, the additive communicational (transmission and reception) energy depletion at each intermediate node makes x264 based intra-coding to be better over MPEG-4 based intra coding.

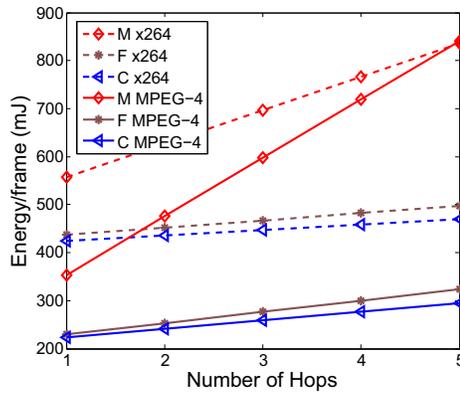
3) *Recommendations Summary*: Based on results presented above, we recommend the following to maximize the lifetime of multi-hop WVSNs. 1) For smaller multi-hop WVSNs, use intra coding whilst for larger multi-hop WVSNs use inter coding; 2) For very small multi-hop WVSNs, use MPEG-4 whilst for larger multi-hop WVSNs use x264; 3) In a medium sized multi-hop WVSNs (e.g., up to 5 hops) with simpler video

contents and/or less stringent requirements of higher video quality, it is recommended to use MPEG-4 based inter coding. However, when the video contents are relatively complex and/or higher video quality is required, it is recommended to use x264 based inter coding.

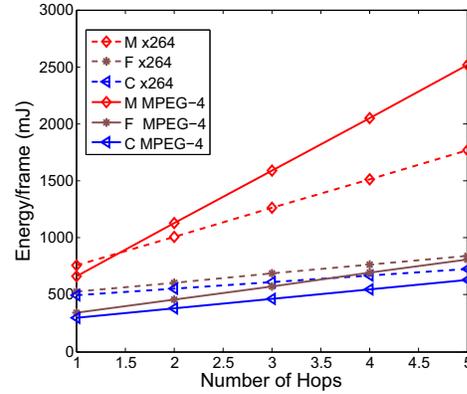
V. CONCLUSION AND FUTURE WORK

In this paper, we empirically evaluated the energy-efficiency for video communication in single-hop and multi-hop WVSNs. We considered different video encoding configurations to exploit the computation-communication tradeoff and find the energy efficient solutions for single-hop and multi-hop WVSNs. Our results show that, in general MPEG-4 video codec is computationally efficient than x264 but worse in terms of compression ratio. For single-hop communication scenario, MPEG-4 based intra coding was observed as the most energy efficient configuration. For multi-hop network with fewer number of hops, inter coding based configurations were generally better over intra coding based configurations. However, with the increase in the number of hops, intra coding based configurations gradually became more energy efficient.

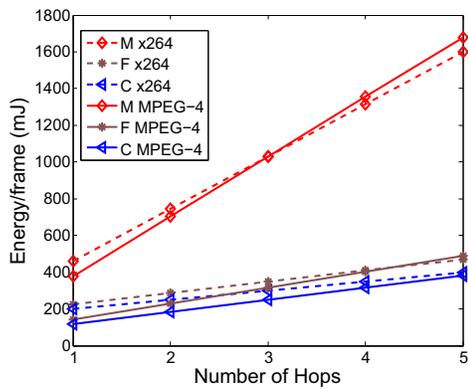
In future, we plan to expand our study to include Distributed Video Coding (DVC) and its variants. DVC is a video coding paradigm which aims to have an asymmetric configuration of simple-encoder and complex decoder. This



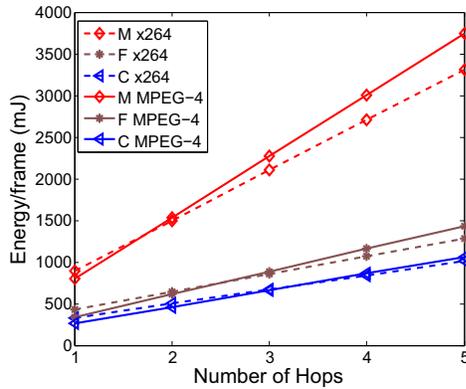
(a) Inter coding at 30dB



(b) Inter coding at 40dB



(c) Intra coding at 30dB



(d) Intra coding at 40dB

Fig. 6. Cross comparison of MPEG-4 and x264 codecs

paradigm ideally suits network configurations with resource-constrained source and powerful destination base stations. For example, WWSNs, mobile phones etc. In addition, we also plan to build an analytical model by jointly considering (computational and communication) overheads, video quality levels, varying network sizes, routing strategies and video compression schemes. Such an analytical model would be able to assist us in analyzing the lifetime maximization problem of WWSNs more concretely.

ACKNOWLEDGMENT

We acknowledge Mr. Adnan Iqbal and Dr. Muhammad Saleem for their valuable comments on the final draft of this paper.

REFERENCES

- [1] I. Akyildiz, T. Melodia, and K. Chowdhury, "A survey on wireless multimedia sensor networks," *Computer Networks*, vol. 51, no. 4, pp. 921–960, 2007.
- [2] T. Kwok and Y. Kwok, "Computation and energy efficient image processing in wireless sensor networks based on reconfigurable computing," in *IEEE Int. Conference on Parallel Processing (ICPP)*, 2006.
- [3] Z. He, W. Cheng, and X. Chen, "Energy minimization of portable video communication devices based on power-rate-distortion optimization," *IEEE Transactions on Circuits and Systems for Video Technology (CSVT)*, vol. 18, no. 5, pp. 596–608, 2008.
- [4] D. Lee, H. Kim, M. Rahimi, D. Estrin, and J. Villasenor, "Energy-efficient image compression for resource-constrained platforms," *IEEE Transactions on Image Processing*, vol. 18, no. 9, pp. 2100–2113, 2009.
- [5] J. J. Ahmad, H. A. Khan, and S. A. Khayam, "Energy efficient video compression for wireless sensor networks," in *43rd IEEE International Conference on Information Sciences and Systems (CISS)*, 2009.
- [6] ISO/IEC, "Information technology – Coding of audio-visual objects – Part 2: Visual," ISO/IEC 14496-2 (MPEG-4), last revised in 2004, 2001.
- [7] I. Richardson, *H. 264 and MPEG-4 video compression*. Wiley Online Library, 2003.
- [8] ITU-T, "Advanced video coding for generic audiovisual services," Rec. H.264, 2003 (last revised in 2007).
- [9] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H. 264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology (CSVT)*, vol. 13, no. 7, 2003.
- [10] "IEEE 802.15.4: Wireless medium access control (MAC) and physical layer (PHY) specifications for low-rate wireless personal area networks (LR-WPANs)," 2006.
- [11] S. Ray, D. Starobinski, and J. Carruthers, "Performance of wireless networks with hidden nodes: a queuing-theoretic analysis," *Computer Communications*, vol. 28, no. 10, pp. 1179–1192, 2005.
- [12] "Crossbow IMote2 Specifications. <<http://www.xbow.com>>."
- [13] "PC based Oscilloscope <<http://www.tiepie.com/uk/home/>>."
- [14] "Open source implementation of MPEG-4 (Part 2), available at <<http://www.xvid.org/Downloads.43.0.html/>>."
- [15] "Open source implementation of H.264 available at <<http://www.videolan.org/developers/x264.html>>."