Video-Aware Rate Adaptation for MIMO WLANs

An Jack Chan*
University of California, Davis, USA
anch@ucdavis.edu

Henrik Lundgren
Technicolor, Paris, France
henrik.lundgren@technicolor.com

Theodoros Salonidis
Technicolor, Paris, France
theodoros.salonidis@technicolor.com

***Abstract*—The IEEE 802.11n standard supports very high physical layer data rates using Multiple Input Multiple Output (MIMO) antenna technologies. Despite such high rates, High Definition (HD) video streaming is still challenging in WLAN deployments. In this paper, we show that the wireless channel probing overhead of existing 802.11n data rate adaptation mechanisms can be detrimental to HD video performance. We propose VARA, a Video-Aware Rate Adaptation protocol that addresses this problem by adapting the frequency and timing of wireless probing to both video encoding rate variations and wireless channel variations. In addition, VARA employs novel strategies that multiplex several Variable Bit Rate (VBR) HD video streams by minimizing their aggregate peak rate requirement. Our experimental evaluations for static and mobile scenarios in a MIMO 802.11n wireless testbed demonstrate the practical benefits of VARA over state-of-the-art 802.11n rate adaptation protocols.**

I. Introduction

Today, video dominates Internet traffic and by 2014 it is projected that nearly half of the video traffic will be 3D or 2D High Definition (HD) [1]. A high fraction of HD video traffic will be consumed by users that access Wireless Local area Networks (WLANs) in homes, enterprises or public spaces. This vision is fueled by two major technology trends. First, recent video streaming technology standards such as H.264/MPEG-4 part 10 AVC [2] reduce HD video bandwidth requirements using Variable Bit Rate (VBR) video encoding. Second, the IEEE 802.11n [3] WLAN standard offers very high wireless physical-layer (PHY) data rates (up to 600 Mbps) using Multiple-Input Multiple-Output (MIMO) antenna technologies.

Despite these advances, the problem of streaming HD video in WLANs is far from being solved. VBR technologies reduce the average video streaming rate by efficient encoding of slow and moderate moving scenes. However, the peak rate remains high as it is determined by the full quality encoding of fast-motion scenes. Also, according to recent studies, the achieved goodput in practical 802.11n deployments can be significantly below the maximum 802.11 PHY data rates [4], [5], due to lack of favorable channel conditions for high MIMO PHY rates, MAC protocol overhead, sub-optimal PHY data rate selection, interference, or even backward compatibility with the previous lower-rate 802.11 standards.

In this paper we show that, in addition and irrespective of the above issues, 802.11n suffers from a fundamental problem that causes video quality degradation. This problem arises due to the probing overhead of the existing 802.11n PHY rate adaptation protocols. These protocols periodically probe different PHY rates to discover the maximum PHY rate supported by the wireless channel. In order to reduce probing overhead, most of these protocols perform *implicit probing* with packets of the ongoing data traffic instead of separate probe packets. Probe packets are lost when transmitted at PHY rates that cannot be supported by the current wireless channel state. Although such probe losses may not be perceptible in delay tolerant data applications, we show that they result in delays or losses of video frames at the application layer, which are both detrimental to HD video streaming performance.

We introduce VARA, a Video-Aware Rate Adaptation protocol, that optimizes wireless channel probing and PHY rate selection by exploiting the VBR streaming rate information of a video. VARA eliminates the channel probing impact on the video stream by scheduling the probes during the low-streaming-rate periods. Furthermore, rather than aggressively trying to find the maximum PHY rate supported by the wireless channel (like all existing 802.11 rate adaptation protocols), VARA selects the most reliable PHY rate that supports the near-future peak streaming rate. To further reduce the probing overhead, VARA monitors the Frame Error Rate (FER) and adapts probing frequency to the measured wireless channel variability.

We then introduce three multiplexing strategies that assist VARA in supporting multiple HD video streams. These strategies inspect and carefully multiplex the different VBR variations of streaming rates in multiple videos to minimize aggregated peak rate, outage time or outage area subject to a capacity target.

Finally, we experimentally evaluate the performance of VARA and the multiplexing strategies using a user-space implementation in a wireless testbed equipped with off-the-shelf 802.11n cards. Our experiments show that VARA reduces the loss rate of higher-rate HD video streams to only a few percent and that this loss reduction translates to 50% video quality increase in terms of Peak Signal-to-Noise Ratio (PSNR). For lower-rate HD video streams, VARA reduces the loss rate to zero and achieves perfect PSNR. During multiple simultaneous video streaming sessions, the multiplexing techniques reduce the aggregated peak rate by up to 25%, which in turn assists VARA to reduce the average loss rate by 20% to 50%.

In summary, our contributions are as follows:

1) We identify the detrimental effect of probing overhead of legacy PHY rate adaptation mechanisms on video

*This work was undertaken while the first author was interning at Technicolor Research Laboratory, in Paris, France.

streaming quality.
2) We introduce VARA, a protocol that adapts PHY rate based on video streaming rate and channel quality.
3) We introduce three multiplexing strategies that assist VARA to efficiently support multiple video streams.

To the best of our knowledge, VARA is the first video-aware wireless rate adaptation protocol for 802.11n MIMO WLANs. In addition to its performance benefits, it is a practical solution that does not require modification of the 802.11 standard. We believe that this approach demonstrates the feasibility and potential of practical cross-layer content-aware techniques in MIMO WLANs and can also be applied to other types of wireless networks.

The rest of the paper is organized as follows. Section II illustrates, explains and quantifies the problem using a practical example in our wireless testbed. In Section III and Section IV, we present the design of VARA and the multiplexing techniques, respectively. In Section V, we experimentally evaluate VARA and the multiplexing strategies in our testbed. Section VII discusses related work and Section VIII concludes.

II. Motivation

In this section we first show an illustrative example where a standard 802.11n rate adaptation protocol results in video streaming quality degradation. We then explain the underlying reasons of the observed problem and finally perform a micro-benchmark to experimentally quantify its impact.

A. Video Streaming Example

We set up a stable 2x3 MIMO 802.11n link in our testbed and measure its maximum capacity[1] over all PHY data rates as 28Mbps. We then stream a video with 26Mbps peak rate over this link with the default automatic rate adaptation protocol (auto rate) option turned on.

Figure 1 shows the video streaming rate and the PHY data rate at the sender and the goodput mesaured at the receiver, during the entire video streaming period. We observe that auto rate most often uses a PHY rate of 52Mbps and periodically probes at 78Mbps. During the peak streaming rate period between 12s and 15s, the goodput drops below the video streaming rate. This indicates that some video packets are lost. Indeed, Figure 2(a) shows that the video packet loss experienced during this period is up to 14% and the video snapshot in Figure 2(b) shows that this loss results in low perceptual video quality.

As an extra validation, we repeat the above experiment with the PHY rate fixed at 52 Mbps, the most frequently selected PHY rate by auto rate during the previous experiment. In this test run with a fixed rate, the video streaming is perfectly supported without any packet loss or quality degradation. This example illustrates that the 802.11n rate adaptation induces a penalty that can significantly degrade the video streaming quality. We proceed to explain the origin of this penalty.

[1]We define as "capacity" of a link at a PHY rate (or auto rate), the maximum (backlogged) UDP goodput measured on this link when this PHY rate (or auto rate) is used.

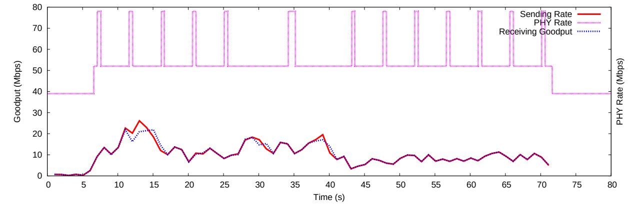
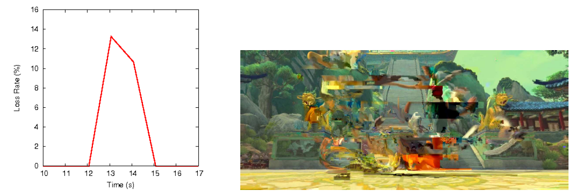


Fig. 1. Video streaming rate, goodput and PHY rate.



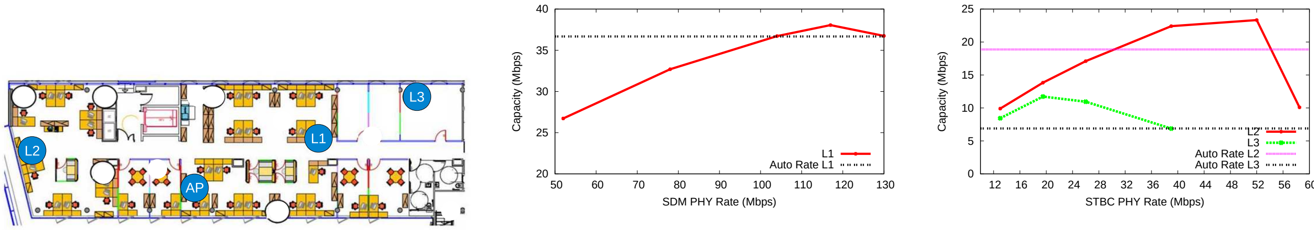
(a) Packet loss (b) Video frame snapshot

Fig. 2. Peak period packet loss and video quality degradation.

B. Probing-induced Capacity Penalty

In order to select an appropriate PHY data rate, 802.11 rate adaptation mechanisms need to probe the wireless channel. Probing can be either explicit or implicit. Explicit probing probes the channel using separate control packets; implicit probing probes the channel using packets from the ongoing data traffic. Most existing ("legacy") rate adaptation protocols use implicit probing due to its reduced overhead. During implicit probing, a few data packets are periodically sent at a higher PHY rate than the current PHY rate. If these transmissions succeed, the channel condition is estimated as adequate to support this higher PHY rate. Then, an even higher PHY rate is used to send the next few data packets. This process continues until a PHY rate where most probe transmissions fail. Then the previous (last successful) PHY rate is selected as the new rate, until the next probing period. The start of the next probing period is triggered by various factors, such as number of successful transmissions, measured packet error rate, time lags and so on, according to different implementations [6].

Increased probing frequency enables faster rate adjustments to channel changes (and thus potentially higher goodput), but also increases the overhead due to failed probe packets. These probing-induced losses result in channel capacity degradation due to binary exponential back-off and airtime consumption for retransmissions. We call it the penalty in capacity utilization. It is important to note that probing and probe losses occur regardless of whether the current channel condition is good or bad. Unless the channel is good enough to support the highest PHY rate all the time, which is rare due to the dynamic property of wireless channels, the rate adaptation will periodically probe the channel to find a better rate to maximize the capacity. A variable channel naturally aggravates the probing overhead as it will cause the probing process to be triggered more frequently and result in increased probing-induced losses.



(a) Access point and node locations (b) Measured capacities at location L1 (c) Measured capacities at locations L2 and L3

Fig. 3. Rate adaptation micro-benchmark testbed and results.

C. Rate Adaptation Micro-benchmark

In order to quantify the probing-induced capacity penalty, we perform a micro-benchmark where we measure the capacity of three different 802.11n links with different fixed PHY rates and auto rate in our testbed, shown in Figure 3(a).

Figure 3(a) shows the location of the access point (AP) and three clients at locations, L1, L2 and L3. For the coherence of our discussion of micro-benchmark, we defer the details of the experimental settings to Section V-A. We use a CBR traffic generator to measure capacities at different fixed PHY rates and auto rate at all three locations. Figures 3(b) and 3(c) show the results. Each data point is the average of 10 4-minute test runs. Instead of a single data point, each auto rate capacity is drawn as a horizontal line for comparison to the fixed PHY rate capacities at the same location.

Figure 3(b) depicts the channel capacities at location L1. At this location, the channel can support the high rate Spatial Division Multiplexing (SDM) mode of MIMO 802.11n where different data streams are sent over different antennas [3]. In this case, higher PHY rates can be used and auto rate achieves close to the maximum capacity of this link. In this case, probing has limited capacity penalty.

Figure 3(c) shows the channel capacities at locations L2 and L3. At these locations the channel conditions deteriorate and SDM mode yields very low or zero goodput. These locations only support the use of the Space Time Block Coding (STBC) mode of MIMO 802.11n, which aims at increasing the robustness by sending copies of a single data stream over different antennas [3]. Figure 3(c) shows that auto rate can only achieve 70%-80% of the maximum capacity. For example, at L2, although auto rate selects PHY rate 52Mbps most of the time, it achieves 75% of the capacity achieved by a fixed PHY rate of 52Mbps.

This micro-benchmark shows that the probing-induced capacity penalty in our testbed ranges between 5% and 25% (and 20% on the average), depending on the channel conditions. Note that the legacy auto rate adaptation mechanism of the 802.11n cards in our testbed employs implicit probing. If explicit probing is used, the probing overhead and capacity penalty would be higher due to the extra control packets required for probing.

We proceed to describe VARA, our video-aware rate adaptation protocol that aims at reducing the probing-induced capacity penalty and minimizing its impact on the video streaming quality by exploiting video streaming information.

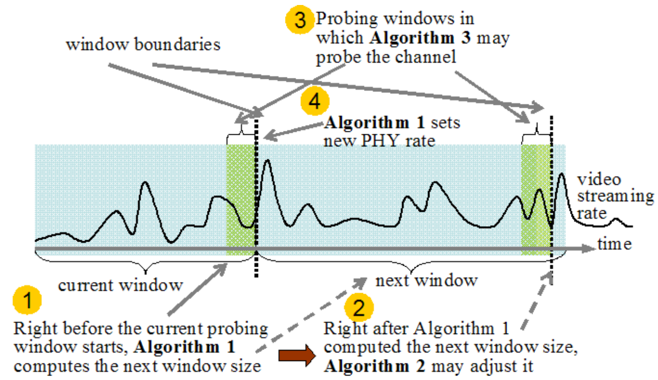


Fig. 4. VARA operation in a video streaming example.

III. VARA DESIGN

VARA is a cross-layer, video-aware PHY rate adaptation protocol. Its basic idea is to use a video streaming rate waveform from the application layer to guide the adaptation of the wireless PHY rate. For any stored video in a wireless home network video server, such a waveform can be easily generated with a play back during video recording.

VARA divides time into variable-sized windows. For each window, VARA attempts to find a PHY rate that yields capacity above the peak video streaming rate in the window. VARA adapts the window sizes to take into account the wireless channel variability and probing overhead.

As shown in the example of Figure 4, VARA consists of three algorithms invoked before the end of each window. First, Algorithm 1 computes the size of the next window based on past measurements of channel variability taken during the current window. Then, Algorithm 2 refines this size to satisfy rate requirements of the probing that Algorithm 3 might run during the next window. If the PHY rate of the current window cannot support the peak video streaming rate of the next window, Algorithm 3 probes the wireless channel for a suitable PHY rate. At the beginning of the next window, Algorithm 1 sets the PHY rate found during the previous steps.

In the next sub-sections, we describe in more detail the operations of these three algorithms.

A. Algorithm 1: window size and PHY rate adaptation

Algorithm 1 is the ”master” algorithm that invokes algorithms 2 and 3. Its operation is depicted in Figure 5. Algorithm 1 is invoked α seconds before the end of the current window. The parameter α is a system-set parameter and is high enough to include the computations and probings described below.

Window size adaptation. Algorithm 1 computes the size of the next window based on the wireless channel variability of the current window. Let n_{total} be the total number of MAC frames (including MAC re-transmissions) transmitted during the current window. The channel variability is computed as the variance $var(\mathbf{L})$ of a set of N Frame Error Rate (FER) values, where $\mathbf{L} = \{l_1, ..., l_N\}$. The i-th FER value l_i, is the fraction of lost MAC frames within the i-th block of c transmitted MAC frames during the current window (i.e., $N = \lfloor n_{total}/c \rfloor$).

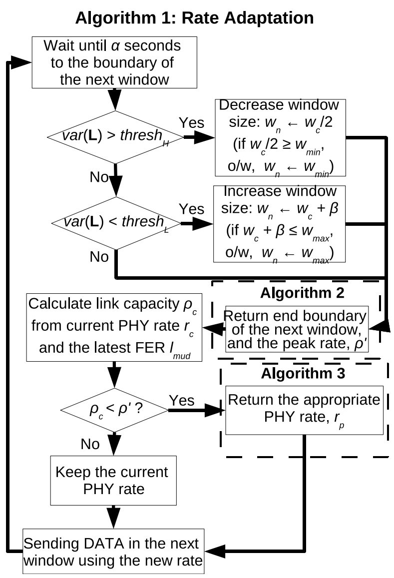


Fig. 5. Flow chart of Algorithm 1.

The size of the next window is determined from the current window size by comparing the variance $var(\mathbf{L})$ to a low threshold $thresh_L$ and a high threshold $thresh_H$. If the variance is lower than $thresh_L$, the current window is additively increased by β seconds; if it is greater than $thresh_H$, the current window is halved; if it is in-between the two thresholds, the current window remains fixed. The window size is bounded by a minimum size, w_{min}, and a maximum size, w_{max}. Multiplicative decrease helps VARA to react rapidly to high channel variability. Additive increase helps VARA slowly discover and lock the window size to a channel variability target.

Algorithm 1 calls Algorithm 2 (Section III-B) to refine and finalize the size of the next window based on the probing overhead requirements. Algorithm 2 returns the refined window boundary and the peak video streaming rate, ρ', of the next window.

Capacity estimation. Algorithm 1 estimates the channel capacity ρ_c of the next window based on the most recently measured FER value l_N and the PHY rate r_c of the current window. Recall that the channel capacity of a PHY rate is defined as the maximum UDP goodput that the PHY rate can achieve in a 802.11n link. It can be computed by the following formula [7]:

$$\rho_c = \frac{P}{t_{idle} + t_{tx}} \tag{1}$$

where P is the UDP payload size used by the video stream, t_{idle} and t_{tx} are the average idle and average transmission times, respectively. The average transmission time t_{tx} is given by:

$$t_{tx} = \frac{P + H}{(1 - l_N)^{ETX} T_{nom}} \tag{2}$$

where H is the UDP header size, and *ETX=1/(1-l_N)* is the expected number of link-layer retransmissions. Also, T_{nom} is the lossless capacity (FER=0), which can be measured offline or computed with analytical formulas [7].

The idle time t_{idle} is given by:

$$t_{idle} = \begin{cases} F(0, \lfloor ETX \rfloor - 1), & \text{if } ETX < m \\ F(0, m-1) + \sigma \frac{(\lfloor ETX \rfloor - m)(W_m - 1)}{2}, & \text{otherwise} \end{cases} \tag{3}$$

where σ is the 802.11 slot duration, $m+1$ is the total number of backoff stages. So, W_0 and W_m are the 802.11 backoff window sizes in stage 0 and in stage m respectively. $F(a, b)$ is the total average backoff time between backoff stages a and b. Assuming the node randomly chooses the backoff window size in a uniform distribution, $F(a, b)$ is the expected backoff window size multiplied by σ.

PHY rate adaptation. Algorithm 1 compares the calculated channel capacity, ρ_c, with the peak streaming rate, ρ', returned by Algorithm 2. If ρ_c exceeds ρ', the PHY rate r_c of the current window will be used in the next window. Otherwise, Algorithm 3 (Section III-C) is called to probe the channel and determine the appropriate PHY rate to use. Once the PHY rate of the next window is determined, Algorithm 1 sets it at the beginning of the next window.

B. Algorithm 2: Window Size Refinement

Algorithm 2 refines the size of the next window computed by Algorithm 1 to handle the probing overhead. Since probing only occurs near the end boundary of the window, the position of the end boundary should be carefully chosen to minimize the impact of probing on video streaming performance.

Let b_n be the end boundary of the next window computed by Algorithm 1. Based on the streaming rate before b_n, Algorithm 2 calculates a probing window size η that can support all probing packets. More specifically, η is computed to satisfy the following:

$$\int_{b_n - \eta}^{b_n} f(t)\, dt > n_p(P_{802.11})(|\,\mathbf{R}\,| - 1) \tag{4}$$

where $f(t)$ is the video streaming rate at time t, n_p is the number of ongoing data frames used for probing at each PHY rate, $P_{802.11}$ is the average WLAN frame size used for video streaming, and **R** is the set of 802.11n PHY rates.[2]

The average video streaming rate ρ_a from time $b_n - \eta$ to b_n should also satisfy:

$$\rho_a < \rho_u \tag{5}$$

The maximum rate requirement ρ_u exists to ensure that the probing overhead will not cause the rate to exceed the peak video streaming rate ρ' of the next window. Such probing overhead causes capacity penalty. From Figures 3(b) and 3(c), the average capacity penalty in our testbed is 20%. Let h be the average capacity penalty caused by probing, then:

$$\rho_u = \frac{1}{1+h}\rho' \tag{6}$$

If (5) is *not* satisfied, Algorithm 2 moves the end boundary of the next window in steps of ξ seconds until it is satisfied. If the current window size was increased by Algorithm 1, Algorithm 2 moves the end boundary later; otherwise it moves it earlier. After each step of moving, b_n, ρ', ρ_u, ρ_a and η are updated accordingly. Once (5) is satisfied, Algorithm 2 returns the final end boundary b_n and the final peak rate ρ'.

As in Algorithm 1, the minimum window size, w_{min}, and the maximum window size, w_{max}, also apply to Algorithm 2. If those limits are exceeded and (5) has not yet been satisfied, Algorithm 2 returns the window size and boundary that was originally computed by Algorithm 1. However, instead of ρ' it returns a higher value $(1+h)\rho'$ for the peak rate. Then, Algorithm 1 will check if the current PHY rate satisfies this higher peak rate which includes the probing overhead. If not, it will call Algorithm 3 to search for a such PHY rate.

C. Algorithm 3: Channel Probing

Algorithm 3 probes the channel when the capacity ρ_c of the current window cannot support the peak streaming rate ρ' of the next window.

During operation, an IEEE 802.11n system must select among 16 PHY rates, that include both SDM and STBC MIMO modes. Algorithm 3 reduces probing overhead by reducing the number of probed PHY rates. This is achieved by using the property that FER is an increasing function of PHY rate within each MIMO 802.11n mode (STBC or SDM) [5]. This in turn implies that, for either STBC or SDM mode, the capacity as a function of PHY rate has a single maximum as shown in Figures 3(b) and 3(c).

Algorithm 3 probes each of SDM and STBC modes separately as follows. First, it determines the probing direction by probing a PHY rate one step lower and a PHY rate one step higher than the current rate r_c. For each rate it uses implicit probing, i.e., it sends n_p consecutive data frames of the ongoing traffic at that PHY rate and measures the FER. Then it uses Equation (1) to compute the corresponding capacity. If both capacities are lower than the capacity ρ_c of r_c, Algorithm 3 returns the rate r_c because it yields maximum capacity for this mode. Otherwise, if the lower (higher) step rate gives higher capacity than ρ_c, Algorithm 3 continues probing all rates at lower (higher) steps one by one until it finds one with a capacity higher than the peak rate requirement ρ'. If no such rate is found, Algorithm 3 yields the PHY rate of maximum capacity.

Finally, Algorithm 3 compares the capacities of the two PHY rates found for SDM and STBC modes and returns to Algorithm 1 the PHY rate whose capacity exceeds ρ' and has a lower FER (i.e., it is more robust). If none of these two capacities exceed ρ' Algorithm 3 returns to Algorithm 1 the PHY rate of higher capacity between the two.

[2]802.11n supports 77 different PHY rates. Offline configurations such as channel width and guard interval reduce them to 16 during network operation [3]. Since the current PHY rate does not require probing, at most $|\mathbf{R}| - 1 = 15$ rates will be probed.

IV. Multiple Video Streams

When multiple video streams exist on a link, VARA treats their aggregated streaming rate as if from a single stream and the same algorithms for handling a single stream can be re-used. However, an aggregate stream of multiple HD videos can have a very high peak rate. A higher peak rate makes it more difficult for VARA to find a matching PHY rate or it may exceed the wireless channel capacity.

VARA addresses this problem using a novel multiplexing technique that we call *Strategic Shifting*. The idea is to strategically shift the starting times of the video stream waveforms such that the peak rate of the aggregate stream is minimized. This technique is enhanced with two other shifting techniques that aim to minimize outage time or outage area with respect to a channel capacity target.

A. Strategic Shifting

Two videos. Suppose a new video session v_2 of duration D_2 is requested to start at time t_q of an ongoing video session v_1 of duration D_1. Strategic Shifting will delay the start of v_2 for δ seconds after t_q. The parameter δ cannot exceed a delay budget of Δ seconds, the highest start-up delay the user of the incoming video can tolerate. Given the two video waveforms $f_{v_1}(t)$ and $f_{v_2}(t)$, the peak aggregate streaming rate $\phi(t_q, \delta)$ when v_2 starts δ seconds after time t_q of v_1 is given by:

$$\phi(t_q,\delta) = \begin{cases} \max\limits_{0\le t\le t_{r_1}} [f_{v_1}(t+t_q+\delta) + f_{v_2}(t)], & t_{r_1} \le D_2 \\ \max\limits_{0\le t\le D_2} [f_{v_1}(t+t_q+\delta) + f_{v_2}(t)], & \text{otherwise} \end{cases} \tag{7}$$

where t_{r_1} is the remaining time $D_1 - (t_q + \delta)$ of video v_1.

Strategic Shifting finds an optimal δ' in $[0, \Delta]$, which yields the minimum aggregated peak rate $\phi(t_q, \delta)$.

$$\delta' = \arg\min_{0\le\delta\le\Delta} \phi(t_q,\delta) \tag{8}$$

Multiple videos. If there are more than two incoming videos, VARA randomly selects one of them to multiplex with the existing video using Strategic Shifting. Then a second video among the incoming videos is randomly selected and multiplexed with the current aggregate stream, and so on. Therefore, the incoming streams are multiplexed sequentially one by one. In this way, Strategic Shifting is repeatedly applied by treating the early admitted videos as a single aggregated stream and multiplexing one new video at a time.

B. Outage Minimized Shifting

Strategic Shifting minimizes peak video streaming rate, but there is still a risk that the channel capacity is exceeded during video streaming. *Outage Minimized (O-M) Shifting* seeks to enhance Strategic Shifting by minimizing the impact of outage when a channel capacity target ρ_c is exceeded. The capacity target parameter can be the average, maximum or most recent capacity measured on the link during the streaming of the existing videos.

If the aggregated peak rate ρ'' returned by Strategic Shifting exceeds ρ_c, O-M Shifting finds a new δ' to minimize *outage time* or *outage area*. The outage time $O_t(\delta)$ is the total time in which the aggregated streaming rate exceeded the channel capacity target ρ_c. The outage area, $O_a(\delta)$ is the area between the waveform of the aggregated streaming rate and the channel capacity target ρ_c. It represents the total number of bits that will be dropped if the target is exceeded and is defined as follows:

$$O_a(\delta) = \int_0^{\max(t_{r_1}, D_2)} [f_{v_1}(t + t_q + \delta) + f_{v_2}(t) - \rho_c] \, dt \quad (9)$$

The performance of both Strategic Shifting and the two versions of O-M Shifting are evaluated in the next section.

V. Evaluation

In this section, we experimentally evaluate VARA's performance using our MIMO 802.11n wireless testbed. We first show that compared to the default auto rate adaptation protocol, VARA significantly reduces packet loss and achieves perfect, or close-to-perfect, video quality in terms of PSNR. We then show that VARA can efficiently adjust windows and schedule probing for different window sizes and video streams. Finally, we show that our multiplexing strategies improve the support for multiple simultaneous video streams.

A. Experimental Settings

We deploy a wireless testbed (Figure 3(a)) in an indoor office environment with cubicles, meeting booths and regular offices. Each testbed node is a Linux PC with Intel Pentium M 1.73GHz processor and 512MB RAM, and is equipped with a Ralink RT2880 802.11n 2T3R MIMO mini-PCI card and three 5dBi omni-directional antennas. We use the RT2880 driver with RT2880iNIC Firmware version 2.0.0.1. The wireless cards are set to operate in channel 108 in the 5GHz frequency band and use their default settings: 20MHz channel width, a Short Guard Interval (SGI) 800ns, and block acknowledgement and frame aggregation features deactivated. The total number of PHY rates including both SDM and STBC modes is 16. i.e., $|\mathbf{R}| = 16$.

One testbed node acts as Access Point (AP) and the others as clients. The AP is at a fixed location while the clients are deployed at different locations with different wireless channel conditions. Comparative experiments are scheduled back-to-back and repeated. To increase experimental repeatability we carry out experiments during evenings and weekends. Also, we use sniffers to ensure that there are no external interferences or hidden terminals during the experiments.

TABLE I
HD movie clips' properties. Rates are in Mbps

Movie Name	Average rate	Peak rate	Variance
Panda1080p	10.26	26.12	28.71
Panda720p	5.96	15.94	10.82
MonsterAliens	5.06	14.64	5.22

B. VARA in Static Environment

We first compare VARA's performance against the default auto rate of the RT2880 Ralink cards (the legacy rate adaptation algorithm) in a static environment. Therefore, we place both the AP and the clients in fixed locations as shown in Figure 3(a). Locations L1, L2, and L3 are selected to yield different wireless channel qualities. L1 has the best, L3 the worst. Table I shows the properties of different HD movie clips used in this experiment. *Panda1080p* represents the high streaming rate video, while *Panda720p* and *MonsterAliens* represent the medium and low streaming rate videos, respectively.

Each experiment run consists of two back-to-back streamings of each HD video between the AP and each client location, first using VARA and then auto rate. We repeat each run five times and show the average.

At the best channel quality location L1 auto rate supports all videos perfectly with zero loss. However, at location L2 it cannot support *Panda1080p*, the highest rate video. In contrast, VARA supports all videos perfectly at both locations L1 and L2. In L3, auto rate cannot support any of the videos perfectly. In contrast, VARA supports all videos perfectly except *Panda1080p*. This is expected because *Panda1080p* has a higher peak rate (26.12Mbps, Table I) than the maximum capacity of L3 (12.5Mbps, Figure 3). As depicted in Figure 6(a), with *Panda1080p* VARA achieves a 2% burst loss lasting for two seconds during the peak rate period, while auto rate results in a burst loss of 35% lasting for six seconds during the peak rate period.

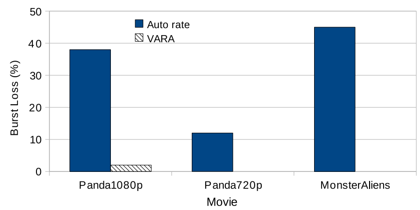
Figure 6(b) shows how these average burst loss rates translate to video quality as measured by average PSNR. We find that even a burst loss rate as small as 12%, as *Panda720p* suffers, can cause a significant degradation of video quality. In terms of subjective video quality, although a video with a PSNR of 25dB to 30dB could still be acceptable, it demonstrates obvious jitters, blocking and blurring. A video with a PSNR around 40dB is considered as a high quality video without any observable defect [8]. We observe that VARA achieves perfect PSNR[3] for two videos, and increases PSNR about 50% for the high rate video over auto rate.

C. VARA in a Mobile Environment

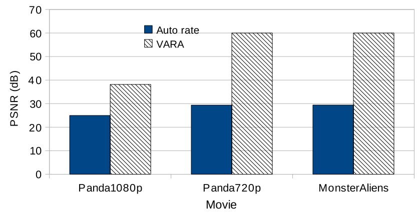
We now evaluate the performance of VARA when the channel condition changes more drastically. We use a controlled mobility scenario where a client moves along the path L4-L5-L6 at walking speed (see Figure 7).

We first measure the FERs and capacities at these locations by sending UDP packets in different PHY rates. Table II shows

[3]By definition, when there is no distortion, PSNR is infinity. For evaluation purposes we use a large number (60dB) to represent such perfect PSNR [9].



(a) Loss rate location L3.



(b) PSNR at location L3.

Fig. 6. Comparing auto-rate and VARA in terms of loss rate and PSNR.

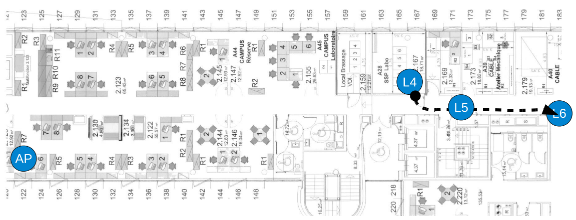


Fig. 7. Testbed mobility experiments.

the averages of these two quantities over five experiment runs[4]. As the client moves from L4 to L5 and then to L6, the FER

TABLE II
AVERAGE FER AND CHANNEL CAPACITY (MBPS) MEASUREMENTS AT LOCATIONS L4, L5 AND L6 FOR DIFFERENT PHY RATES. CAPACITIES ARE IN BOLD.

Data Rate	19.5	26	39	52
L4	-	-	0, **30.34**	98.5, **0**
L5	-	0.12, **19.24**	0.45, **10.27**	0.79, **0.02**
L6	0, **16.75**	0.49, **5.89**	0.59, **3.92**	-

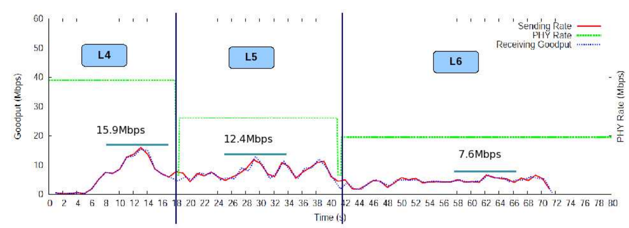
of a particular PHY rate changes. For example, when 39Mbps PHY rate is used and when the client moves to L4, L5 and to L6, the FER increases from 0, to 45% and to 59%, respectively. At the same time, the capacity decreases.

We then perform an experiment in which a client moves with the same mobility pattern using VARA and auto rate while the *Panda720p* HD movie clip is being streamed. Figure 8(a) shows the impact of VARA's rate adaptation on the goodput in this scenario. VARA computes different window sizes when the client is in different locations. Window size increases as the variation of FER is small when the current PHY rate is used. (It is worth noting that a large FER does not necessary mean a large FER variation.) When the end boundary of each window approaches, VARA evaluates if the current PHY data rate can provide the capacity large enough for the peak rate of the next window.

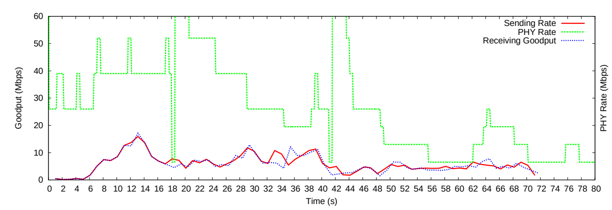
[4]The dashes exist because we only measured at PHY rates that VARA selected in subsequent video streaming experiment to save experiment time.

By comparing the peak rate requirements denoted in Figure 8(a) at the windows and the capacities in Table II, VARA selects the appropriate PHY rates. At L4, when the first window starts, VARA selects PHY rate of 39Mbps, whose capacity of 30.34Mbps can satisfy the peak rate of 15.9Mbps. When the end boundary of the first window approaches, VARA discovers that the capacity of PHY rate 39Mbps cannot satisfy the peak rate of 12.4Mbps of the next window as the client has already moved to L5. By comparing the capacities yielded by 52Mbps and 26Mps, it finds that 26Mbps can support the peak rate. It then selects 26Mbps and uses it in the next window. Similarly, when the client moves to L6, VARA selects a PHY rate of 19.5Mbps for the next window.

Figure 8(a) also shows that VARA ensures that the goodput curve always matches the video streaming rate curve. An exception are the time instants where rate changes and the PHY rate shows deep spikes. These spikes are due to our user-space implementation. When the user-space issues a PHY rate change to the driver, there is a short-term interruption which causes the capacity reduction. Such effects can be removed with a kernel space VARA implementation.



(a) Video streamings when VARA is used.

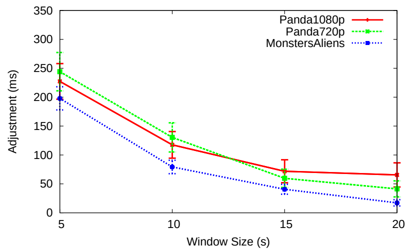


(b) Video streaming when the auto rate is used.

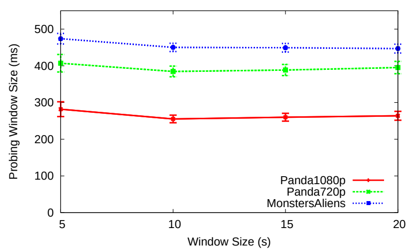
Fig. 8. Video streaming in a mobile environment.

Figure 8(b) shows the streaming rate, goodput and PHY rate changes when auto rate is used in the same scenario. Auto rate also tries to select the right PHY rate when the client moves from L4, to L5 and to L6. However, frequent unnecessary probings consume the capacity especially when the channel condition degrades (e.g., in L5 and L6). We can deduce the lack of channel capacity from the non-overlapping of the sending rate and the goodput curves. In addition, there is 30% burst loss rate when auto rate is used while there is no loss when VARA is used.

Furthermore, from our traces we find that the lack of capacity caused by the frequent unnecessary probings also increases the end-to-end delay of the streaming to around 500ms. In contrast, when VARA is used, the end-to-end delay is within 10ms.



(a) Adjustment of the end boundary.



(b) Probing window size.

Fig. 9. Effect of window size.

D. Analysis of window adjustments and probing windows

The results of both static and mobile experiments validate that VARA selects the right PHY rates and can significantly improve the video streaming quality. In this section, we use the three videos in Table I to evaluate the ability of Algorithm 2 to compute the end boundary and the probing window size η of the next window so that video streaming rate requirements are satisfied and probing overhead is minimized.

Window adjustments. To find the right PHY rates, VARA still needs probings. Therefore, it is crucial that Algorithm 2 in VARA can successfully find the appropriate boundaries of windows such that the probing will not lead to video streaming quality degradation. Even though an appropriate boundary is found, if the adjusted boundary computed by Algorithm 2 deviates from the original boundary computed by Algorithm 1 too much, it will drastically change the window size computed by Algorithm 1. Since Algorithm 1 in VARA computes the window size to adapt to the channel condition change, a large boundary adjustment by Algorithm 2 will lead to ineffectiveness in responding to the channel condition change.

Figure 9(a) shows that, as the window size increases, the adjustment by Algorithm 2 decreases. In VARA, probing occurs near the end boundary of the window. Therefore, the end boundary should not be set in the period in which the streaming rate is high. When the window size increases, due to the rate variability in VBR videos, there is a higher chance that the streaming rate near the original boundary is not close to the peak rate in the window. Therefore, as the window size increases, the adjustment needed by Algorithm 2 decreases.

Probing window size. In VARA, the probing process must be completed before the current window ends. If the probing window is too large, it will also lead to the decrease of the probability of finding an appropriate end boundary.

Figure 9(b) shows the probing window size computed by Algorithm 2 as a function of window size, for $n_p = 10$ probes, and also $|R| = 16$ PHY rates. The probing window size does not strongly depend on the window size. We also observe that the probing window size for panda1080p is the smallest while that of MonsterAliens is the largest among the three. This is because the probing window size is determined by the streaming rate. With implicit probing, videos with high streaming rate can finish the probing process faster than those with low streaming rate, since during the probing process a high streaming rate can send more probes per time unit.

Success rate. In all cases, the success rate for finding the appropriate boundary is over 95%. It is worth noting that when Algorithm 2 cannot find an appropriate end boundary, it does not mean VARA fails. In that case, Algorithm 2 will return a rate equal to $(1 + h)\rho'$ as the peak rate. Then Algorithm 1 will check if the current PHY rate can provide a capacity large enough for that returned peak rate, or Algorithm 3 will be called to find a such PHY rate.

The results of this section further prove the effectiveness of VARA. In all cases, the adjustments of the end boundary are small (less than 0.5s). Such small adjustments by Algorithm 2 would not change the window size computed by Algorithm 1 too much. That means the original window size of Algorithm 1 which adapts to the channel condition change is largely preserved. Moreover, the small probing window size (all are less than 0.5s) implies that even if VARA sets small windows to adapt to the dynamic channel condition change, it is still easy to find the period within the window for channel probing.

E. Reducing the Peak Rate by Strategic Shifting

We now investigate the ability of Strategic Shifting to minimize the aggregated peak rate of multiple videos in practice. To find the optimal shift δ' that minimizes the aggregated peak rate, our implementation uses exhaustive search to solve Equation (8). The computation delay and the quality of the solution of this implementation depend on the granularity of δ, the delay budget Δ, and the length of the videos.

We test the algorithm on four video clips whose lengths range from 72s to 2.5 minutes and their peak rates range from 14.54Mbps to 26.12Mbps. Three of them are from the videos in Table I. We also add another lower-rate HD video, *Wolfman*, which has the similar rate properties as *MonsterAliens* in this evaluation. We first use a granularity of 0.2s for δ and test the algorithm with two delay budgets $\Delta = 5s$ and $\Delta = 10s$, respectively.

At each experiment, we start a video and within the first five seconds we randomly schedule a time where one, two, or three other videos request to start streaming simultaneously.

In all experiments, our implementation of Strategic Shifting multiplexed any two streams within one second. Figure 10 shows the resulting aggregated peak streaming rates. We observe that Strategic Shifting provides an aggregated peak rate reduction between 15% and 25% compared to the case where no shift is applied. Thus, a δ granularity of 0.2s can already

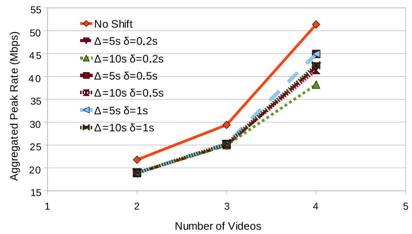
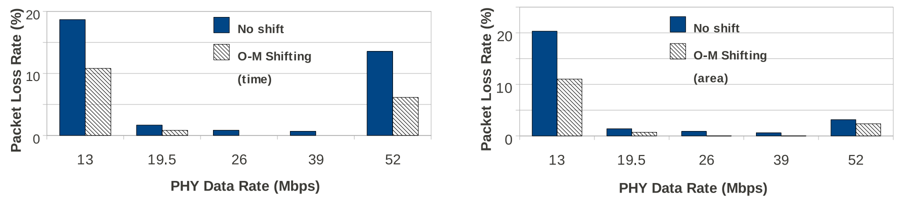


Fig. 10. The aggregated peak streaming rate when different number of videos are multiplexed.



(a) The outage time is minimized

(b) The outage area is minimized

Fig. 11. The video streaming loss rate when using different O-M shifting.

demonstrate the benefits of Strategic Shifting. A δ granularity less than 0.2s might reduce the aggregated peak rate further, but it would also incur higher computation delay. In Figure 10, we also present the results when larger granularities, 0.5s and 1s, are used. They also demonstrate significant peak rate reduction compared to the case where no shift is applied. Furthermore, a larger Δ allows Strategic Shifting to further reduce peak rate, especially as the number of videos increases, as shown in 4 video case in Figure 10.

F. Reducing Loss by Outage Minimized (O-M) Shifting

We now evaluate the performance of Outage Minimized (O-M) Shifting in practice. Each experiment consists of two back-to-back runs, where two 2.5-minute HD videos (*Wolfman* and *MonsterAlien*) are streamed together on a link. The experiments are carried out when the client is in the location L3 in Figure 3(a).

In the first run we multiplex the two videos using O-M Shifting with $\Delta = 5s$ and δ search granularity of 0.2s. In the second run, no shifting is applied. Each experiment is repeated ten times and performed at different PHY rates.

Figure 11 shows the average packet loss rate of the second video stream.[5] Both O-M Shifting strategies of minimizing outage time and minimizing outage area, result in 20% to 50% reduction in packet loss rates compared to the no-shift case. At some PHY rates, which achieve the capacity of the wireless link, O-M Shifting removes all losses and supports the video streaming perfectly.

Similar to Strategic Shifting, if we use a larger delay budget, Δ, and a smaller granularity for the shift, δ, O-M Shifting can reduce the loss rate further, at the cost of higher computation time.

VI. Discussion

Our VARA evaluations focused on basic scenarios where one or more video flows are streamed on a single link. VARA can also operate in more complex scenarios such as hidden terminals, multiple competing links or when video traffic co-exists with best-effort traffic such as file downloads.

In presence of hidden terminals (and high collision rates), VARA will measure a higher FER and make a conservative decision of a low PHY rate. This is a common issue in all rate adaptation mechanisms and can be addressed by methods of separating collisions from channel errors (e.g. [7]). When multiple links compete, each link will select a PHY rate based on its video rate requirement and its own estimate of remaining channel capacity. When best-effort traffic is present, VARA can operate transparently with existing traffic differentiation mechanisms which would give higher priority to the video traffic (e.g. 802.11e traffic classes).

In addition, VARA selects the minimum rate that satisfies the video rate requirement. On the one hand, this results in more robust transmissions than higher eligible PHY rates. On the other hand, it results in packets of longer duration which may be more susceptible to interference or slow down other competing links transmitting at higher PHY rates.

In our future work we plan to perform a comprehensive evaluation of VARA in the above scenarios.

VII. Related Work

802.11 rate adaptation mechanisms. References [6], [10]–[12] do not probe at multiple rates but make rate adaptation decisions based on SNR [10], [11], Channel State Information (CSI) [12] or FER [6] measurements at the current rate and then increase or decrease the rate based on pre-computed thresholds and rate lookup tables. However, lookup tables and thresholds change dynamically with the wireless channel and require in-situ training [13]. Also, most 802.11 chipsets cannot accurately measure SNR due to interference [14] and, except the implementation of [12] on Intel's iwl5300 chipsets, 802.11n chipsets typically do not export CSI from PHY layer to firmware or driver level. SampleRate [15] probes the channel by selecting a different PHY rate at random at every tenth data packet and reduces probing overhead by avoiding PHY rates that yield low link quality. In [5], Pefkianakis et al. introduce rate adaptation techniques for MIMO 802.11n systems which reduce implicit probing overhead by exploiting the monotonicity property between FER and PHY rate within SDM or STBC MIMO mode. SoftRate [16] adapts the PHY rate by differentiating whether the frame loss is due to the collision or channel error. However, it requires PHY layer information not provided by standard 802.11 hardware and shows significant throughput improvement only in mobile scenarios after training.

VARA differs from the above approaches in several aspects. First, it is the first approach that takes into account video rate requirements. Second, it is fully compatible with the

[5]The loss rate of the first video is similar to that of the second video.

802.11 standard. Third, it measures FER and uses probing thus avoiding the challenge of determining rate tables and thresholds in a dynamic environment. Fourth, it not only reduces the probing frequency and probing overhead as [5] but also takes a cross-layer approach and uses video streaming information to efficiently schedule the probing during low-streaming-rate periods. This enables VARA to eliminate the probing's negative impact on video streaming. Finally, the techniques of separating collisions from channel errors developed in [16], [17] and in [7] are orthogonal to VARA and can assist on making more accurate rate adaptation decisions.

Video streaming systems. Recently, different systems are used for video streaming. Adaptive streaming techniques (e.g. HTTP adaptive streaming) rely on a server hosting multiple copies of a video in different streaming rates and qualities. Scalable Video Coding (SVC) encodes a video into different layers [18]. Chunks from the different copies or layers can be periodically requested to adapt the streaming rate to changing network capacities. The availability of multiple copies is typically limited to large-scale online video services with large storage capacities. For SVC, the clients should also be equipped with special SVC decoders. VARA does not rely on such systems. In addition, VARA takes an orthogonal approach and optimizes the wireless rates to best accommodate the video stream, rather than the other way around.

VBR smoothing. There is a body of work that considers smoothing VBR video streaming by decoupling the VBR video streaming rate from the actual transmission rate (see e.g., [19] and references therein). These schemes typically use *a priori* knowledge about the video stream to schedule pre-fetching of packets to the buffer to lower streaming peak rate and variance. Our strategic shifting aims at reducing peak streaming rate in the case of multiple streams. Our work differs in that it does not change the transmission scheduling from the server. We also do not need a large buffer in the receiver side for storing the pre-fetched packets. Our work multiplexes several video streams, whereas smoothing schemes typically consider only a single video. Moreover, recent work [20] has shown that VBR traffic characteristics after smoothing exhibits significant variability with H.264 compared to older encoding standards used in [19]. Our Strategic Shifting is therefore complementary to VBR smoothing.

VIII. Conclusions

Supporting HD VBR video in WLANs is a timely and challenging problem. Although WLAN technologies such as 802.11n MIMO support very high wireless PHY rates, we showed that in practice the probing overhead of existing state of the art 802.11n PHY rate adaptation protocols can be detrimental to video performance.

Our wireless rate adaptation protocol (VARA) addresses this problem by adapting the frequency and timing of wireless probing to the video streaming rate and the wireless channel variations. We also proposed three novel Shifting techniques to efficiently multiplex HD videos by minimizing peak aggregate streaming rate, outage time and outage area.

The feasibility of VARA was demonstrated in an indoor 802.11n testbed. Our experiments showed that VARA and the Shifting techniques can efficiently support one or more HD video streams. Compared to the default auto rate adaptation protocol it can eliminate or reduce packet loss rates to a few percents. For the video quality in PSNR, VARA can increase the lower-rate HD video to original perfect quality, while it can increase the PSNR of the higher-rate HD video by 50%.

To the best of our knowledge, VARA is the first video-aware wireless rate adaptation protocol to support HD videos in 802.11n WLANs. In our future work we will seek to further optimize HD video performance by combining VARA's wireless rate adaptation approach with adaptive streaming approaches such as HTTP streaming and SVC.

References

[1] Cisco System, "Cisco visual network index: Forecast and methodology, 2009 - 2014," *Cisco While Paper*, 2010.

[2] ISO/IEC 1449610:2003, "Coding of Audiovisual ObjectsPart 10: Advanced Video Coding, also ITU-T Recommendation H.264 Advanced video coding for generic audiovisual services." 2003.

[3] IEEE 802.11 Working Group, "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications – Amendment 5: Enhancements for Higher Throughputs," 2009.

[4] K. Pelechrinis, T. Salonidis, H. Lundgren, and N. Vaidya, "Experimental Characterization of 802.11n Link Quality at High Rates," in *Proc. ACM WiNTECH*, 2010.

[5] I. Pefkianakis, Y. Hu, S. H. Wong, H. Yang, and S. Lu, "MIMO Rate Adaptation in 802.11n Wireless Networks," in *Proc. ACM MobiCom*, 2010.

[6] S. H. Y. Wong, H. Yang, S. Lu, and V. Bharghavan, "Robust Rate Adaptation for 802.11 Wireless Networks," in *Proc. ACM MobiCom*, 2006.

[7] T. Salonidis, G. Sotiropoulos, R. Guerin, and R. Govindan, "Online Optimization of 802.11 Mesh Networks," in *Proc. ACM CoNEXT*, 2009.

[8] Y. Wang, "Survey of objective video quality measurements," in *Technical report*. Worcester Polytechnic Institute, 2006.

[9] A. Chan, K. Zeng, P. Mohapatra, S.-J. Lee, and S. Banerjee, "Metrics for Evaluating Video Streaming Quality in Lossy IEEE 802.11 Wireless Networks," in *Proc. IEEE INFOCOM*, 2010.

[10] G. Holland, N. Vaidya, and P. Bahl, "A rate-adaptive mac protocol for multi-hop wireless networks," in *Proc. ACM MobiCom*, 2001.

[11] G. Judd, X. Wang, and P. Steenkiste, "Efficient channel-aware rate adaptation in dynamic environments," in *Proc. ACM MobiSys*, 2008.

[12] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Predictable 802.11 packet delivery from wireless channel measurement," in *Proc. ACM SIGCOMM*, 2010.

[13] J. Camp and E. Knightly, "Modulation rate adaptation in urban and vehicular environments: Cross-layer implementation and experimental evaluation," in *Proc. ACM MobiCom*, 2008.

[14] J. Zhang, K. Tan, J. Zhao, H. Wu, and Y. Zhang, "A practical snr-guided rate adaptation." in *Proc. IEEE INFOCOM*, 2008.

[15] J. Bicket, "Bit-rate selection in wireless networks," in *Master's Thesis*. MIT, 2005.

[16] M. Vutukuru, H. Balakrishnan, and K. Jamieson, "Cross-layer wireless bit rate adaptation," in *Proc. ACM SIGCOMM*, 2009.

[17] S. Rayanchu, A. Mishra, D. Agrawal, S. Saha, and S. Banerjee, "Diagnosing wireless packet losses in 802.11: Separating collision from weak signal," in *Proc. IEEE INFOCOM*, 2008.

[18] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the h.264/avc standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, 2007.

[19] J. Salehi, Z.-L. Zhang, J. Kurose, and D. Towsley, "Supporting stored video: reducing rate variability and end-to-end resource requirements through optimal smoothing," *IEEE Transactions on Networking*, 1998.

[20] G. Van der Auwera, P. David, M. Reisslein, and L. Karam, "Traffic and quality characterization of the h.264/avc scalable video coding extension," *Journal Advances in Multimedia*, 2008.