

Joint Voice/Video Retry Limit Adaptation for On-Demand Streaming over WiFi Networks

Author 1, Author 2, Author 3, and Author 4

Abstract—This letter proposes a fast retry limit adaptation algorithm for 802.11e distributed networks. Differently from existing solutions, which manage a unique access category, the presented method jointly estimates the retry limits associated to both voice and video packets. The algorithm, which operates in saturated and non-saturated traffic conditions, is validated adopting an 802.11n physical layer and comparing its performance with a proper extension of an existing solution.

Index Terms—Wireless audio/video streaming, distributed network, 802.11e, retry limit, distortion.

I. INTRODUCTION

A recent study released by the European Commission has revealed that Wireless Fidelity (WiFi) is the most widespread technology for streaming applications, and that this hegemony is expected to further grow with the diffusion of Gigabit-WiFi and WiFi-enabled smartphones [1]. To manage the audio/video contents in a distributed network, the 802.11e enhanced distributed channel access (EDCA) introduces four access categories (ACs) of decreasing priority: voice (VO), video (VI), best effort (BE), and background (BK) [2]. The parameters that characterize these ACs, such as the arbitration inter-frame space (AIFS) and the minimum contention window, are specified by the EDCA according to the adopted physical (PHY) layer. Instead, the maximum number of retransmissions (retry limit) is not subject to mandatory specifications. Thus, it has been exploited by several studies to control the perceived quality in contention-based scenarios [3–7]. These studies provide effective retry limit estimations considering machine learning techniques [3], virtual buffers [4], relative priority indexing [5], content-aware strategies [6], and stochastic models [7]. The common aspect of these proposals is that the adaptation involves a single AC. However, in a distributed network, the VO/VI ACs may be both active [8], since some users may have to jointly exchange audio and video streams on a peer-to-peer basis for WiFi offloading purposes, or a mesh router may have to simultaneously receive audio and video contents from other nearby mesh routers. A joint VO/VI retry limit adaptation may hence represent a desirable advance to support the coexistence and limit the reciprocal degradation of independent audio and video flows in these scenarios.

Accordingly, this letter presents a fast algorithm for on-demand WiFi streaming that jointly estimates the retry limits of VO/VI packets according to the distortion introduced by their possible loss in saturated and non-saturated conditions.

The letter is organized as follow. Section II formulates the problem. Section III presents the algorithm. Section IV discusses the results and summarizes the main conclusions.

II. SYSTEM MODEL

Consider an 802.11e distributed network with N contending sources and denote the considered ACs by $q = 1$ (VO) and $q = 2$ (VI). Besides, as in [4], assume a Poisson packet arrival process of average rate λ , which can also model the saturated traffic scenario for large average rates [5–7].

Each source sends an audio sequence \mathcal{S}_1 encoded by the G.729 standard to obtain a set $\{s_1^k : k = 1, \dots, K_1\}$ of K_1 packets [9], and a video sequence \mathcal{S}_2 encoded by the H.264 standard to obtain a set $\{s_2^k : k = 1, \dots, K_2\}$ of K_2 packets [10]. The quality of a sequence \mathcal{S}_q^k , decoded in absence of a lost packet s_q^k [6], is estimated as $\mathcal{Q}_q^k = \mathcal{Q}(\mathcal{S}_q, \mathcal{S}_q^k)$ by using, as quality assessment measure $\mathcal{Q}(\cdot, \cdot)$, the perceptual evaluation of speech quality (PESQ) if $q = 1$ [11], and the cumulative structural similarity (SSIM) if $q = 2$ [12]. The unity-based normalized distortion associated to each packet s_q^k for $k = 1, \dots, K_q$ and $q = 1, 2$ is hence evaluated as:

$$D_q^k = 1 - (\mathcal{Q}_q^k - a_q)/(A_q - a_q), \quad (1)$$

where $a_q = \min\{\mathcal{Q}_q^1, \dots, \mathcal{Q}_q^{K_q}\}$ and $A_q = \max\{\mathcal{Q}_q^1, \dots, \mathcal{Q}_q^{K_q}\}$. Since the letter's objective is the retry limit adaptation given a distortion measure, a common single-loss offline estimation of \mathcal{Q}_q^k is assumed. This complies with an on-demand scenario. However, the proposed adaptation accepts as input any \mathcal{Q}_q^k value, obtained accounting for multiple losses or evaluated through a fast online distortion estimation algorithm [13].

The sources adopt the EDCA basic access, thus the average success time T is identical to the collision time. For the VO/VI ACs, the maximum backoff stages are $m_1' = m_2' = 1$ and the AIFS values are equal [2], thus the two ACs differ just for the minimum contention window W_q . Hence, during the access procedure, the reactivation of a backoff counter previously frozen because of channel occupation requires that the channel is sensed idle for an identical time, regardless that the packet belongs to a VO or VI AC. Besides, the impact of the BE/BK ACs on the VO/VI ones is limited [7]. These characteristics imply that the VO and VI ACs may be described by two Markov chains with the same structure and without transitions among them. Thus, the analysis may be inferred from the single-AC model in [14]. To avoid cumbersome repetitions, the entire mathematical derivation is not reported, rather focusing on the basic equations that will be involved in the development of the adaptation algorithm.

The core of the model is a nonlinear system that provides, for $q = 1, 2$, the conditional collision probability p_q , the transmission probability τ_q , the packet arrival probability during the processing of a previous packet $\eta_{q,1}$, and the packet arrival

probability when the source is idle $\eta_{q,2}$. According to [7], [14], this system may be expressed in the following form:

$$\tau_q = \left[W_q + \frac{1}{2} + \left(\frac{1 - \eta_{q,1}}{\eta_{q,2}} - \frac{W_q}{2} \right) \frac{1 - p_q}{1 - p_q^{m_q+1}} \right]^{-1} \quad q=1, 2 \quad (2)$$

$$p_q = 1 - (1 - \tau_1)^{N-2+q} (1 - \tau_2)^{N-1} \quad q=1, 2 \quad (3)$$

$$\eta_{q,1} = 1 - e^{-\lambda S(p_q)E(\tau_1, \tau_2)} \quad q=1, 2 \quad (4)$$

$$\eta_{q,2} = 1 - e^{-\lambda E(\tau_1, \tau_2)} \quad q=1, 2 \quad (5)$$

where m_q is the (unknown) retry limit,

$$S(p_q) = \left(W_q - \frac{1}{2} \right) \frac{1 - p_q^{m_q+1}}{1 - p_q} - \frac{W_q}{2} \quad (6)$$

is the average number of backoff counter decrements, and:

$$E(\tau_1, \tau_2) = T - (T - \nu)(1 - \tau_1)^N (1 - \tau_2)^N \quad (7)$$

is the average time required for a backoff counter decrement when a slot time ν is adopted. In particular, (2), which generalizes [14, eqs.(27),(34-I)] to two ACs for a unity maximum backoff stage, states that a transmission occurs when the backoff counter becomes equal to zero. Besides, (3), which directly derives from [7, eq.(9-II)] for $q = 1, 2$, models the collisions, which occur if, at the beginning of the same slot, transmissions are attempted by at least two sources (external collision), or by two ACs of the same source (internal collision). In this second case, the VO packet is transmitted and the VI packet is considered collided [2]. The last two equations, which generalize [14, eqs.(34-III,IV)] to two ACs, express the probabilities that a packet arrives at the transmission queue of the source. More precisely, (4) accounts for the time $S(p_q)E(\tau_1, \tau_2)$ spent for the access procedure of a previous packet, while (5) accounts for the time $E(\tau_1, \tau_2)$ spent for waiting a packet at an empty queue. Further mathematical details for the derivation of (2)-(7) may be found in [7], [14].

The p_q and D_q^k values play a fundamental role in retry limit adaptation problems. The aim is to guarantee a higher (lower) retry limit to a packet whose loss produces a higher (lower) distortion [3]. Furthermore, to account for the network load, the retry limit should be higher (lower) when the probability of not colliding $1 - p_q$ gets lower (higher). However, as suggested in [5], the retry limit should not be directly proportional to $(1 - p_q)^{-1}$, but rather to its order of magnitude, in order to avoid the generation of huge and hence not applicable values when p_q becomes close to one. Accordingly, the retry limits for $k = 1, \dots, K_q$ and $q = 1, 2$ may be evaluated from:

$$\hat{m}_q^k = \operatorname{argmin}_{m_q \in \mathbb{N}} |m_q - \alpha_q D_q^k - \beta_q \log(1 - p_q)^{-1}|, \quad (8)$$

where α_q and β_q are positive weights introduced to manage the distortion and the collision effects. The addressed problem hence becomes that of finding \hat{m}_q^k by solving (2)-(8). The solution of this problem and the description of the resulting adaptation algorithm are the objectives of the next section.

III. ALGORITHM

As a first step, p_1 and p_2 are derived from (2)-(7) by distinguishing between non-saturated and saturated traffic.

A. Non-saturated traffic

In the non-saturated case, the arrival probabilities are strictly lower than one. Therefore, one may adopt the approximation $e^x \cong 1 + x$ in (4) and (5), thus obtaining:

$$\eta_{q,1} \cong \lambda S(p_q)E(\tau_1, \tau_2), \quad (9a)$$

$$\eta_{q,2} \cong \lambda E(\tau_1, \tau_2), \quad (9b)$$

for $q = 1, 2$. Substituting (9) and (6) in (2), one obtains:

$$\tau_q \cong \left[1 + \frac{1}{\lambda E(\tau_1, \tau_2)} \frac{1 - p_q}{1 - p_q^{m_q+1}} \right]^{-1}, \quad (10)$$

for $q = 1, 2$. According to the non-saturation hypothesis, one may assume $p_1, p_2 \ll 1$ and low m_q values, which imply $(1 - p_q)/(1 - p_q^{m_q+1}) \cong 1$ in (10) and hence $\tau_2 \cong \tau_1$. Applying these approximations and substituting (7) in (10), one obtains, after some calculations, the following algebraic equation:

$$\pi_\gamma(t) = t^\gamma - \frac{\lambda T + 1}{\lambda(T - \nu)} t + \frac{1}{\lambda(T - \nu)} = 0, \quad (11)$$

of degree $\gamma = 2N + 1$ in the unknown $t = 1 - \tau_1$. Since γ is odd, $\pi_\gamma(t)$ has at most three real roots. Besides, $\pi_\gamma(0) > 0$ and $\pi_\gamma(1) < 0$, since $T > \nu$, and $\pi_\gamma(t) \rightarrow \pm\infty$ for $t \rightarrow \pm\infty$. Therefore, $\pi_\gamma(t)$ has exactly three real roots, with a unique one \bar{t} in the interval $[0, 1]$. This latter root enables the calculation of the conditional collision probabilities in the non-saturated case. In fact, recalling (3) for $1 - \tau_2 \cong 1 - \tau_1 \cong t$, yields:

$$p_q = 1 - \bar{t}^{2N-3+q}, \quad q=1, 2. \quad (12)$$

B. Saturated traffic

In the saturated case, the arrival probabilities become very close to one and the m_q values get higher, thus $\eta_{q,h} \cong 1$ for $q, h = 1, 2$, and $(1 - p_q)/(1 - p_q^{m_q+1}) \cong (1 - p_q)$. Using these approximations in (2) and substituting (3) in the resulting expressions, one may derive the pair of equations:

$$t_q = 1 - 2 \left[W_q \left(1 + t_1^{N-2+q} t_2^{N-1} \right) + 1 \right]^{-1}, \quad q=1, 2, \quad (13)$$

in the unknowns $t_q = 1 - \tau_q$ for $q = 1, 2$. The equation corresponding to $q=1$ in (13) may be solved for t_2 as:

$$t_2 = \frac{1}{t_1} \left[\frac{1 + t_1 + 2W_1(t_1 - 1)}{W_1(t_1 - 1)} \right]^{\frac{1}{N-1}}. \quad (14)$$

Now, substituting (14) in (13) for $q=2$, and remembering that $W_2 = 2W_1$ [2], one may obtain, after some manipulations, the algebraic equation:

$$\pi_\delta(t_1) = (\varepsilon_{00} - 1)(t_1 - 1) \left[2\varepsilon_{10}t_1^3 - (\varepsilon_{31} - 2)t_1^2 + \varepsilon_{21}t_1 \right]^{N-1} - (\varepsilon_{10}t_1 - \varepsilon_{11})(2\varepsilon_{10}t_1^2 - \varepsilon_{31}t_1 + \varepsilon_{20})^{N-1} = 0, \quad (15)$$

of degree $\delta = 3N - 2$, where the terms $\varepsilon_{ij} = 2^i W_1 + (-1)^j$ for $i = 0, \dots, 3$ and $j = 1, 2$ are introduced to obtain a compact representation of (15). Regardless of whether δ be odd or even, $d\pi_\delta(t_1)/dt_1 < 0$ for $t_1 \in [0, 1]$ (calculations are cumbersome and are omitted). Moreover, $\pi_\delta(0) > 0$ and $\pi_\delta(1) < 0$, since $W_1 > 1$. Hence, $\pi_\delta(t_1)$ has a unique root \bar{t}_1 in the interval $[0, 1]$. Also in this case, this specific root allows the derivation

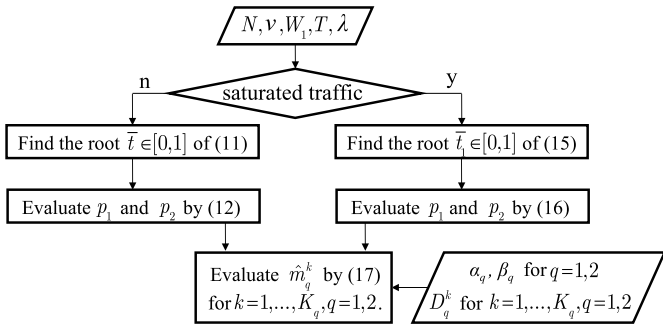


Fig. 1. Proposed algorithm.

of p_q for $q = 1, 2$. In fact, remembering that $\tau_q = 1 - t_q$ for $q = 1, 2$, and using (14) in (3), one may evaluate the conditional collision probabilities in the saturated case as:

$$p_q = 1 - 2\bar{t}_1^{q-1} + \frac{\bar{t}_1^{q-1}(1 + \bar{t}_1)}{W_1(1 - \bar{t}_1)}, \quad q = 1, 2. \quad (16)$$

C. Retry limit estimation

Once the p_1 and p_2 values are derived, the retry limit of the generic packet s_q^k , for $k = 1, \dots, K_q$ and $q = 1, 2$, may be immediately calculated from (8) as:

$$\hat{m}_q^k = \lfloor \alpha_q D_q^k - \beta_q \log(1 - p_q) \rfloor, \quad (17)$$

where $\lfloor \cdot \rfloor$ denotes the round function. All elements necessary to elaborate the adaptation algorithm are now available.

The algorithm develops as follows (Fig. 1). Consider a network with N sources and characterized by the quantities ν, W_1, T, λ , which are identified by the incoming traffic and the selected PHY layer. As a first step, find the root $\bar{t} \in [0, 1]$ of (11) and then evaluate (12) (non-saturated case), or find the root $\bar{t}_1 \in [0, 1]$ of (15) and evaluate (16) (saturated case). As a second step, given the weights α_q and β_q , calculate \hat{m}_q^k according to D_q^k and p_q by (17) for $k = 1, \dots, K_q$ and $q = 1, 2$.

The first advantage of this algorithm is its ability of jointly estimating the VO/VI retry limits in saturated and non-saturated conditions. The second advantage is its simplicity, since (11) and (15) may be quickly solved by efficient root-finding methods, and (12), (16), (17) are available in closed-form. This feature is emphasized by the fact that, for a given network scenario, a unique p_1/p_2 estimation is sufficient for all VO/VI packets, because (11), (12), (15), (16) are independent of k . Moreover, the weak dependence of (17) from moderate changes of p_q due to data rate variations, makes the algorithm applicable to adaptive PHY rate scenarios.

IV. RESULTS AND CONCLUSIONS

The algorithm is validated considering independent audio and video sequences and using an 802.11n PHY layer [15] (Table I), with an average success/collision time [2,15]:

$$T = \text{AIFS} + H/\rho_c + \Lambda/\rho + \text{SIFS} + \text{ACK}/\rho_c. \quad (18)$$

The first set of presented results aims to check the suitability of the adopted approximations and to clarify how the algorithm

S_1	<i>The Barber of Seville</i>	SIFS	16 μ s
S_2	<i>Foreman</i>	AIFS	34 μ
Data rate ρ	120 Mbps	ν	9 μ s
Control rate ρ_c	24 Mbps	W_1	4
Average payload Λ	300 bytes	T	82.67 μ s
Header DATA H	24 bytes	α_q	qN for $q = 1, 2$
ACK	14 bytes	β_q	q for $q = 1, 2$

TABLE I
ADOPTED SEQUENCES AND PARAMETERS.

operates. Fig. 2 shows the distortion of the VO sequence (Fig. 2(a)), and the corresponding retry limits obtained in three network scenarios that differ for the N and λ values (Fig. 2(b)). These results, which are limited to the first 50 packets of the VO sequence for readability reasons, reveal the accuracy of the introduced approximations. In fact, for each scenario, the estimated retry limits (markers), which are evaluated by the proposed algorithm considering non-saturated traffic for $\lambda = 10^2$ p/s and saturated traffic for $\lambda = 10^4$ p/s, properly match the theoretical retry limits (lines), which are evaluated in the presence of BE/BK saturated traffic by numerically solving (2)-(8) for each packet, that is, using the success/collision time corresponding to the actual payload.

Concerning the algorithm's behavior, Fig. 2 shows that, as desired, each sequence of retry limits follows the trend of the distortion. The weight α_q , which manages the impact of the distortion, controls the range of the \hat{m}_q^k values, while the weight β_q , which manages the impact of the collision probability, controls the minimum \hat{m}_q^k value. Exploiting these characteristics, the settings $\alpha_q = qN$ and $\beta_q = q$ for $q = 1, 2$ (Table I) have been adopted to obtain a finer estimation when the number of sources increases and/or the priority decreases.

Fig. 3 reports the SSIM and PESQ (in the legend) obtained in the three scenarios by running 20 network simulations

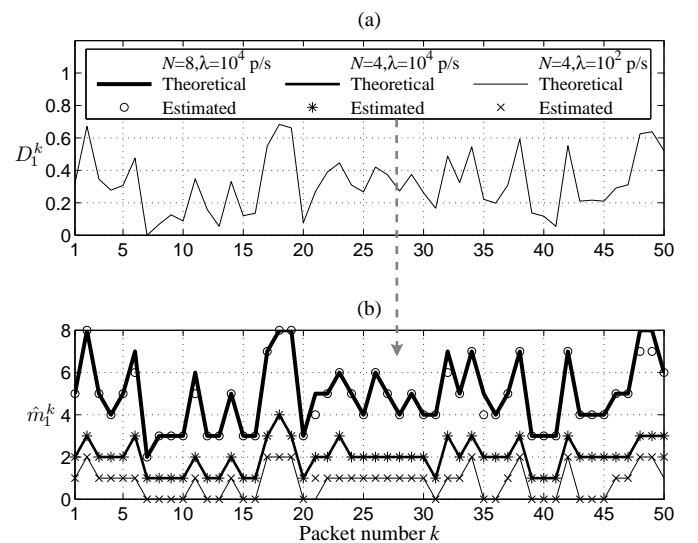


Fig. 2. Distortion (a) and theoretical and estimated retry limits (b) for the first 50 packets of S_1 considering different values of N and λ (in packets/s).

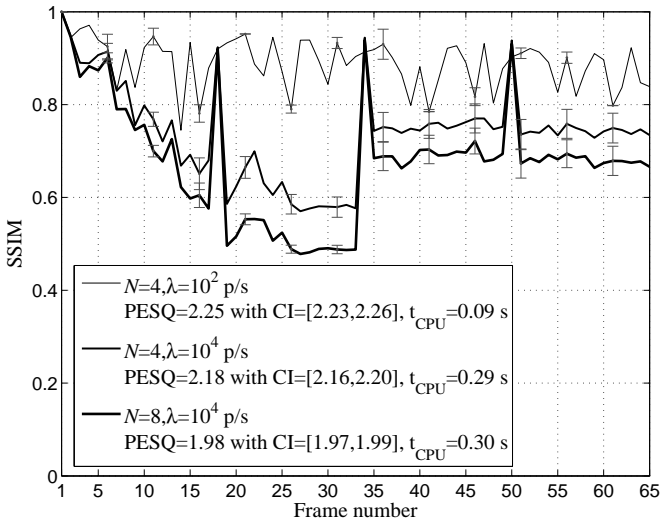


Fig. 3. SSIM, PESQ, and CPU time for different values of N and λ .

of 10 s each (sufficient to accomplish the access procedure for all VO/VI packets of all nodes), and then averaging the results over the simulations and the nodes. The simulations are performed on a Matlab 802.11e/n packet-level simulator implementing the EDCA as a state machine and running on one core of an Intel Core2 Quad Q9300 @2.50 GHz Sun Ultra 24 workstation. The figure also reports the 95% confidence intervals (CIs) and the CPU time t_{CPU} necessary to estimate all the \hat{m}_q^k values for a given scenario. These values confirm that the algorithm is very fast, since the entire estimation process always required less than half second.

A further set of simulations is carried out to obtain the PESQ in Fig. 4(a) and the SSIM (averaged also over frames) in Fig. 4(b) as a function of N in saturated conditions. Since, to the best of authors' knowledge, alternative joint VO/VI retry limit adaptations are not currently available, a basis for comparison has been obtained by extending the scheme in [5]. In particular, [5] presents a classification-based method where \bar{k}_2 low-priority VI packets are associated to a lower retry limit r , dependent on the collision probability, and the other $K_2 - \bar{k}_2$ high-priority VI packets are associated to a higher retry limit R . The aim is to guarantee that the high-priority packets experience half of the loss probability compared to the low-priority ones. Following this approach, one may derive a lower retry limit r_q for each VO/VI AC by expressing [5, eq.(4)] in compact form and then generalizing it as:

$$r_q = \lfloor R + \log 2 / \log p_q \rfloor, \quad q=1, 2, \quad (19)$$

where p_q , which is not estimated in [5], is evaluated by (16). Furthermore, to obtain a fair comparison for each sequence S_q , the total number of retry limits estimated by the proposed algorithm $\sum_{k=1}^{K_q} \hat{m}_q^k$, is set equal to that estimated by the extension of [5] $\bar{k}_q r_q + (K_q - \bar{k}_q)R$. This equality provides the number of low-priority packets associated to r_q :

$$\bar{k}_q = \left\lfloor \frac{K_q R - \sum_{k=1}^{K_q} \hat{m}_q^k}{R - r_q} \right\rfloor, \quad q=1, 2. \quad (20)$$

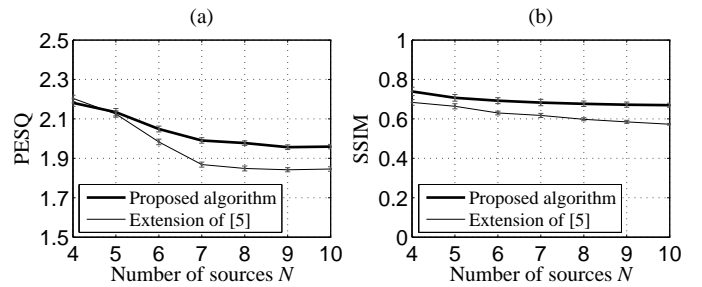


Fig. 4. PESQ (a) and SSIM (b) as a function of N in saturated conditions.

Applying these settings for $R = N$, one obtains the retry limits producing the thin curves in Fig. 4. Both approaches required a CPU time of 0.30 s. The figure reveals that the finer retry limit estimation of the proposed algorithm (Fig. 2), provides a higher performance as compared to that achievable by the classification-based approach, which, for a given AC, has to select each retry limit among just two values r_q and R .

In conclusion, a fast retry limit adaptation algorithm for 802.11e distributed networks able to jointly manage the VO/VI ACs in saturated and non-saturated traffic conditions has been presented. The results have shown that the proposed algorithm provides the advantages of a fine retry limit estimation maintaining a low CPU time. The conceived solution may be hence specifically suited to improve the quality of on-demand streaming over the increasingly widespread WiFi networks.

REFERENCES

- [1] J.S. Marcus and J. Burns, "Study on Impact of Traffic Off-Loading and Related Technological Trends on the Demand for Wireless Broadband Spectrum," EUC DG Content, Tech. Rep., 2013.
- [2] *IEEE WLAN MAC/PHY Specif.: QoS Enhanc.*, IEEE Std 802.11e, 2005.
- [3] M. van der Schaar, D.S. Turaga, and R. Wong, "Classification-Based System For Cross-Layer Optimized Wireless Video Transmission," *IEEE Trans. Multimedia*, vol. 8, no. 5, pp. 1082–1095, Oct. 2006.
- [4] H. Bobarshad, M. van der Schaar, A.H. Aghvami, R.S. Dilmaghani, and M.R. Shikh-Bahaei, "Analytical Modeling for Delay-Sensitive Video Over WLAN," *IEEE Trans. Multimedia*, vol. 14, no. 2, pp. 401–414, Apr. 2012.
- [5] Y. Zhang, Z. Ni, C.H. Foh, and J. Cai, "Retry Limit Based ULP for Scalable Video Transmission over IEEE 802.11e WLANs," *IEEE Commun. Lett.*, vol. 11, no. 6, pp. 498–500, Jun. 2007.
- [6] C.-M. Chen, C.-W. Lin, and Y.-C. Chen, "Cross-Layer Packet Retry Limit Adaptation for Video Transport over Wireless LANs," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1448–1461, Nov. 2010.
- [7] F. Babich, M. Comisso, and R. Corrado, "Fast Retry Limit Adaptation for Video Distortion/Delay Control in IEEE 802.11e Distributed Networks," *Ad Hoc Netw.*, vol. 36 (Part 1), pp. 229–243, Jan. 2016.
- [8] Z. Wan, N. Xiong, N. Ghani, M. Peng, A.V. Vasilakos, and L. Zhou, "Adaptive Scheduling for Wireless Video Transmission in High-Speed Networks," in *IEEE INFOCOM HSN Work.*, 2011.
- [9] *Speech Coding at 8 kbit/s Using CS-ACELP*, ITU-T Rec. G.729, 1996.
- [10] *AVC for Generic Audiovisual Services: SVC*, ITU-T Rec. H.264, 2012.
- [11] *PESQ: E2E Speech Quality Assessment*, ITU-T Rec. P.862, 2001.
- [12] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [13] M. Schier and M. Welzl, "Optimizing Selective ARQ for H.264 Live Streaming: A Novel Method for Predicting Loss-Impact in Real Time," *IEEE Trans. Multimedia*, vol. 14, no. 2, pp. 415–430, Apr. 2012.
- [14] F. Babich and M. Comisso, "Throughput and Delay Analysis of 802.11-Based Wireless Networks Using Smart and Directional Antennas," *IEEE Trans. Commun.*, vol. 57, no. 5, pp. 1413–1423, May 2009.
- [15] *IEEE WLAN MAC/PHY Specif.: HT Enhanc.*, IEEE Std 802.11n, 2009.