

Analysis of an Adaptive Modulation and Coding Scheme with HARQ for TCP Traffic

Onur Ozturk

Department of Electrical and
Electronics Engineering
Bilkent University
Ankara, Turkey 06800
Email: ozturk@ee.bilkent.edu.tr

Nail Akar

Department of Electrical and
Electronics Engineering
Bilkent University
Ankara, Turkey 06800
Email: akar@ee.bilkent.edu.tr

Abstract—In this paper, we analyze the aggregate TCP throughput performance of a wireless link utilizing Active Queue Management (AQM) and an Adaptive Modulation and Coding (AMC) scheme with Hybrid ARQ (HARQ) based on the probability of failure in the first transmission attempt. We assume packets arriving out-of-order at the wireless receiver due to random retransmissions are resequenced before being released to the network. For this reason, an approximate model for the delay experienced at the resequencing buffer is also presented. In the light of the results obtained from the presented analysis, we propose a threshold for the aforementioned probability of failure making the investigated AMC scheme work at an overall performance close to that of the optimum policy.

Keywords—TCP, Adaptive Modulation and Coding, HARQ

I. INTRODUCTION

Adaptive Modulation and Coding (AMC) refers to the class of algorithms in mobile communication systems used to select one of the modulation and coding schemes (MCSs) offered by the system's air interface so as to satisfy certain Quality of Service (QoS) requirements without having to sacrifice from spectral efficiency. In particular, MCSs are generally indexed for increasing spectral efficiency and AMC tracks channel state in real-time in order to switch to the MCS resulting in acceptable levels of performance in terms of throughput, packet error rate (PER), delay, jitter, etc. for a given QoS class. Applications relying on transport control protocol (TCP) are generally delay insensitive but loss intolerant. TCP itself, however, is both loss and delay tolerant which allows optimization of the loss and delay experienced to maximize its throughput. In this regard, Hybrid ARQ (HARQ), for which lost packets at the receiver are retransmitted and information from all (re)transmissions is combined to enhance forward error correction (FEC) performance, matches very well with the basic TCP operation. However, TCP triggers its error recovery and congestion control mechanism for out-of-order packet reception, which in turn significantly drops its throughput. HARQ, on the other hand, re-orders its packets in the course of retransmissions. Re-ordered packet arrivals are reacted in the same manner with packet losses yielding severe degradation in TCP throughput. It is reported in [1] that TCP throughput of a single flow may drop approximately by 90% provided that 10% of the packets are re-ordered in three locations. There is therefore a need to restore the original packet order at

the wireless receiver by means of a resequencing mechanism whose drawback is increased round-trip-times (RTTs) which might adversely affect TCP throughput.

In [2], the authors derive closed form analytical expressions for various performance metrics of different HARQ schemes, but they do not particularly study the TCP protocol. Reference [3] analyzes TCP performance of HARQ with Active Queue Management (AQM) but assumes ACK/NACK feedback for retransmissions to be instantaneous. In a more recent study [4], the authors analytically compare the performances of HARQ and ARQ schemes for TCP but they take neither packet re-ordering nor AQM into account. In [5], $M(x)/G/1$ queuing system of [6] is adopted to relate workload (queue occupancy level) dependent loss and delay parameters of a wireless AQM router to aggregate TCP level throughput with the ultimate aim of evaluating performance of a Traffic Agnostic Link Adaptation (TAGLA) scheme with single transmission opportunity. TAGLA, indifferent to any TCP layer parameter, makes a selection among the offered MCSs by its Physical layer (PHY) based on their individual capacities and PER statistics. In this paper, we generalize the framework presented in [5] to accommodate the HARQ transmission technique. From a modeling perspective, one needs to address the following issues raised by the use of HARQ: (i) the workload increase caused by retransmissions, (ii) enhanced PHY decoding performance reflected to the resulting PER, (iii) retransmission delays and (iv) out-of-order packet arrivals at the TCP receiver. We address the effect of out-of-order packet arrivals by claiming resequencing to be an indispensable and complementary part of any HARQ-TCP system which guarantees in-order packet delivery at the expense of an additional resequencing delay. In the current paper, we address these issues on the basis of the queuing model of [5]. Since the workload-dependent queuing framework of [5] is extensively validated for a wide variety of traffic scenarios, in terms of both packet delay and loss values, this effort is not replicated in this paper. On the other hand, the proposed approximate model for the resequencing delay is validated with MATLAB[®] simulations and is shown to lead to an acceptable level of accuracy. The presented analytical method is then used to evaluate performance of a TAGLA scheme with HARQ (TAGLAWH) tracking the probability of failure in the first transmission attempt. We note that with HARQ, higher rates of packet loss can be targeted for the first transmission attempt compared to the single transmission

case, which renders such an algorithm highly feasible even for channels with high variability. As in [5], we take IEEE 802.16 [7] as the underlying PHY technology, but the presented analysis can be used with other technologies.

The rest of the paper is organized as follows. In Section II, we present the workload-dependent queuing model along with an approximate resequencing delay model for HARQ. In Section III, first the proposed resequencing delay model is validated and then the numerical results for the performance of the TAGLAWH scheme are presented. We conclude in the final section.

II. ANALYTICAL MODEL

In this paper, we analyze a wireless link which is the bottleneck link (i.e. packet losses and bandwidth limitations at other links are negligible) for a fixed number, say N , of contending TCP flows. The link is offered with M MCSs denoted by MCS_m , where $m \in [0, M-1]$, operating at an SNR level SNR_s , where $s \in [0, S-1]$, among S possible SNR levels. PHY of the system transmits each packet at a raw bit rate of r_m bps by means of FEC blocks of k_m bytes. Upon failure of decoding a FEC block at the receiver, the block is retransmitted up to Z times adhering to the Selective Repeat policy using Type-I HARQ with Chase Combining decoding (HARQ-CC). The TCP flows have arbitrary but fixed round-trip-times denoted by $RTT_{0,i}$ for flow i , $i \in [0, N-1]$. Packets have a common and deterministic packet length of L bytes and are subject to AQM at the ingress of the wireless transmitter's First-In-First-Out (FIFO) queue. The packets experience a two-way framing delay $2D_F$ and a mean resequencing delay $X_{m,s}$ accounting for the delay caused by both the random retransmissions and the resequencing process at the receiver. TCP ACK packets transmitted at the reverse path are assumed to be prioritized so that associated delay and losses can be neglected.

A. $M(x)/G/1$ Queuing Model

We adopt the description and the accompanying notation for workload-dependent $M/G/1$ queues from [6]. In this paper, we model the wireless link as an $M(x)/G/1$ queue with Poisson TCP packet arrivals with a workload-dependent intensity function $\lambda(x)$, and a deterministic workload-dependent service rate $r(x)$ (in units of bps), where x is the instantaneous workload (in units of bits) of the queue. Packets contribute to the workload of the queue by a job size (in units of bits) whose Cumulative Distribution Function (CDF) is denoted by $B(\cdot)$. It is shown in [6] that the steady-state workload density $v(\cdot)$ satisfies the following integro-differential equation for $x > 0$:

$$r(x)v(x) = \lambda(0)V(0)(1 - B(x)) + \int_{y=0^+}^x (1 - B(x-y))\lambda(y)v(y) dy, \quad (1)$$

with a non-zero atom $V(0)$ at $x = 0$, for finite length buffer space [5].

In order to find $\lambda(x)$ in (1), we use the so-called PFTK TCP model of [8] which relates the packet loss rate and RTT seen by a flow to its TCP throughput. Let p , λ , and T_0 denote the packet loss rate, packet send rate, and the TCP retransmission timeout parameter of a TCP source, respectively. In our model,

we use the following relationship used in the implementation of TCP in [9]:

$$T_0 = \max(T_{0,\min}, RTT + 4\sigma_{RTT}), \quad (2)$$

where RTT and σ_{RTT} are the estimates for the mean and standard deviation of RTT, respectively, and $T_{0,\min}$ is the minimum value the timeout parameter can take. Let W_u and $W_{\max} = W/L$ denote the random variables associated with the unconstrained window size and the maximum window size (in units of packets) of the TCP source, where W is the TCP receiver's buffer size. Also let b denote the number of packets referred by a single cumulative ACK packet sent by the TCP receiver. The reference [8] proposes the following equation to relate the TCP packet send rate λ to p and RTT seen by a flow:

$$\lambda = \begin{cases} \frac{\frac{1-p}{p} + E[W_u] + \tilde{Q}(E[W_u])\frac{1}{1-p}}{RTT(\frac{b}{2}E[W_u] + 1) + \tilde{Q}(E[W_u])T_0\frac{f(p)}{1-p}}, & E[W_u] < W_{\max} \\ \frac{\frac{1-p}{p} + W_{\max} + \tilde{Q}(W_{\max})\frac{1}{1-p}}{RTT(\frac{b}{8}W_{\max} + \frac{1-p}{pW_{\max}} + 2) + \tilde{Q}(W_{\max})T_0\frac{f(p)}{1-p}}, & \text{otherwise,} \end{cases} \quad (3)$$

where

$$f(p) = 1 + p + 2p^2 + 4p^3 + 8p^4 + 16p^5 + 32p^6, \quad (4)$$

$$\tilde{Q}(w) = \min\left(1, \frac{(1 - (1-p)^3)(1 + (1-p)^3)(1 - (1-p)^{(w-3)})}{1 - (1-p)^w}\right), \quad (5)$$

and

$$E[W_u] = \frac{2+b}{3b} + \sqrt{\frac{8(1-p)}{3bp} + \left(\frac{2+b}{3b}\right)^2}. \quad (6)$$

Throughout our numerical studies, we fix $T_{0,\min} = 0.2s$ as in [10], $b = 2$, $W = 64$ Kbytes as in [9] and $L = 1500$ bytes. With these, (3) provides a closed-form expression for the TCP send rate λ in terms of p , RTT and σ_{RTT} .

Loss component p in (3) is assumed to be comprised of the wireless packet errors and the intentional packet drops of the AQM policy which are statistically independent from each other. We assume Gentle variant of RED, GRED [11], to be the selected AQM scheme for regulating TCP traffic whose drop policy is as follows:

$$p(x) = \begin{cases} 0, & 0 \leq x < th_{\min} \\ \frac{x - th_{\min}}{th_{\max} - th_{\min}} p_{\max}, & th_{\min} \leq x < th_{\max} \\ p_{\max} + \frac{x - th_{\max}}{th_{\max}} (1 - p_{\max}), & th_{\max} \leq x < 2th_{\max} \\ 1, & \text{otherwise.} \end{cases} \quad (7)$$

In this paper, th_{\min} and th_{\max} are set to $30L$ and $90L$, respectively, in units of bits, and p_{\max} is set to 0.1 as in [5]. Let $FER_{m,s,z}$ denote the FEC block error rate at the z^{th} retransmission, where $z \in [0, Z]$. Then overall wireless packet loss probability $PER_{m,s}$ can be found as

$$PER_{m,s} = 1 - \left(1 - \prod_{z=0}^Z FER_{m,s,z}\right)^{F_m}, \quad (8)$$

where $F_m = \lceil L/k_m \rceil$ stands for the minimum number of FEC blocks required to transmit a single packet, yielding an effective packet length of $L_m = F_m k_m$ bytes.

The RTT term of each flow also consists of multiple components as follows:

$$RTT_i(x) = RTT_{0,i} + 2D_F + L_m/r_m + x/r_m + X_{m,s}, \quad (9)$$

where L_m/r_m and x/r_m are the transmission and the queuing delays, respectively. We let V_m and R_m be the modulation order and the code rate of MCS_m and further decompose $r(x) = r_m$ as in (10), where r is the symbol rate of PHY.

$$r_m = r \log_2(V_m)R_m. \quad (10)$$

Furthermore, let $T_{0,i}(x)$ denote the workload-dependent timeout parameter for flow i which can be expressed via (2) as

$$T_{0,i}(x) = \max(T_{0,\min}, RTT_i(x) + 4\sigma_{RTT,i}), \quad (11)$$

where $\sigma_{RTT,i}$ stands for the standard deviation of RTT for flow i . The term $\sigma_{RTT,i}$ can be related to the standard deviation of the queuing delay $\sigma_{Q_{m,i}}$ and that of the resequencing delay $\sigma_{X_{m,s}}$ as follows:

$$\sigma_{RTT,i} = \sqrt{\sigma_{X_{m,s}}^2 + \sigma_{Q_{m,i}}^2}. \quad (12)$$

We refer reader to [5] for the formulation of $\sigma_{Q_{m,i}}$ and derive resequencing statistics $X_{m,s}$ and $\sigma_{X_{m,s}}$ in the next section.

The overall rate of packets that are admitted into the queue denoted by $\lambda(x)$ can then be written as

$$\lambda(x) = \sum_{i=0}^{N-1} (1 - p(x))\lambda_i(x), \quad (13)$$

where $\lambda_i(x)$ is the send rate of flow i when the queue occupancy takes the value x . We propose to use the PFTK TCP model (3) to write $\lambda_i(x)$ by replacing RTT and T_0 with their per-flow workload-dependent counterparts $RTT_i(x)$ and $T_{0,i}(x)$, respectively. In a similar fashion, p is replaced by \bar{p}_i , the average packet loss probability for flow i , as in (14),

$$\bar{p}_i = 1 - (1 - PER_{m,s})(1 - \bar{q}_i), \quad (14)$$

where \bar{q}_i denotes the queue average packet loss probability stemming only from AQM for flow i .

Errored packets in the first transmission attempt are retransmitted using HARQ-CC for which all retransmitted packets are identical. Retransmissions are assumed to be made in a Selective Repeat manner for which only the packets received in error are retransmitted after a retransmission delay of D_R unless Z number of retransmission opportunities are exhausted for each packet. PMF $f_{\mathbf{H}_{m,s}}(h)$ for the random variable (RV) $\mathbf{H}_{m,s}$ denoting the number of retransmissions of each FEC block can be expressed as follows:

$$f_{\mathbf{H}_{m,s}}(h) = \begin{cases} 1 - FER_{m,s,0}, & h = 0 \\ (1 - FER_{m,s,h}) \prod_{z=0}^{h-1} FER_{m,s,z}, & 0 < h < Z \\ \prod_{z=0}^{Z-1} FER_{m,s,z}, & h = Z. \end{cases} \quad (15)$$

Assuming loss events of building FEC blocks of a packet to be i.i.d Bernoulli distributed, total number of retransmissions required for each packet, denoted by the RV $\mathbf{G}_{m,s}$ is the maximum of those of the F_m FEC blocks as shown in (16).

$$\mathbf{G}_{m,s} = \max_{F_m} \{\mathbf{H}_{m,s}\} \quad (16)$$

Retransmissions increase the workload of the transmitter which can be modeled as a virtual increase in the packet length. We let $f_{\mathbf{A}_{m,s}}(x)$ be the PDF of the virtual packet length $\mathbf{A}_{m,s}$ and relate it to PMF $f_{\mathbf{G}_{m,s}}(g)$ of $\mathbf{G}_{m,s}$ as follows:

$$f_{\mathbf{A}_{m,s}}(x) = \sum_{g=0}^Z f_{\mathbf{G}_{m,s}}(g)\delta(x - (g+1)L_m), \quad (17)$$

where $\delta(\cdot)$ is the dirac delta function. Note that IEEE 802.16 PHY [7] requires whole packet to be retransmitted even if a single FEC block is errored at the receiver which is captured by equation (17). Integrating $f_{\mathbf{A}_{m,s}}(x)$, we find $B_{m,s}(x)$ as in (18).

$$B_{m,s}(x) = \int_{y=0}^x f_{\mathbf{A}}(y) dy \quad (18)$$

Finally, we let $P_{m,s,z}$ denote the probability of failure of the z^{th} retransmission and derive it from $f_{\mathbf{G}_{m,s}}(z)$ as follows:

$$P_{m,s,z} = \begin{cases} 1 - f_{\mathbf{G}_{m,s}}(0), & z = 0 \\ 1 - \frac{f_{\mathbf{G}_{m,s}}(z)}{\prod_{t=0}^{z-1} P_{m,s,t}}, & 0 < z < Z \\ \frac{PER_{m,s}}{\prod_{t=0}^{Z-1} P_{m,s,t}}, & z = Z. \end{cases} \quad (19)$$

As we present all components of (1), we refer reader to [5] for the remaining steps to reach a complete solution. We finalize this section by providing a method to numerically solve the integro-differential equation (1). For that purpose, queue occupancy is discretized with a discretization interval Δ such that $L/\Delta \gg 1$ is an integer. We then define $v_i = v(i\Delta)$, $\lambda_i = \lambda(i\Delta)$ and $B_i = B(i\Delta)$ for $i > 0$ and discretize (1) to obtain

$$v_i = \begin{cases} \frac{\lambda_0 V(0)(1-B_i) + \sum_{j=1}^{i-1} (1-B_{i-j})\lambda_j v_j \Delta}{r - \lambda_i \Delta}, & i < l \\ \frac{\sum_{j=i-l+1}^{i-1} (1-B_{i-j})\lambda_j v_j \Delta}{r - \lambda_i \Delta}, & \text{otherwise,} \end{cases} \quad (20)$$

where l is the integer such that $B(l\Delta) = 1$. Note that the identity (20) enables the calculation of v_i as a weighted sum of v_j 's for $j < i$. We propose to set $V(0) = 1$ and iteratively calculate v_i for $1 \leq i \leq K$ as in (20), where $\sum_{i=1}^{\infty} v_i \Delta = 0$. Throughout the paper we safely set $K = (2th_{\max} + (Z+1)L_m)/\Delta$ and $\Delta = 20$ bits. We define $V = V(0) + \sum_{i=1}^K v_i \Delta$ and then normalize the quantities $V(0)$ and v_i as follows:

$$V(0) := V(0)/V, v_i := v_i/V, \quad 1 \leq i \leq K. \quad (21)$$

B. Approximate Resequencing Model

The receiver maintains a resequencing buffer to re-order packets arriving out-of-order due to random retransmissions. The presumed resequencer in this analysis waits for the successful decoding of missing packets in its buffer and either upon their arrival or expiration of their timeouts imposed by the hard limit Z , releases all subsequent packets delayed for the arrival of these packets to the network. We note that the described resequencer is generic, since it does not adhere to a particular implementation. In order to find the mean resequencing delay denoted by $X_{m,s}$, we first derive the associated PDF by taking similar steps with [12]. Briefly, reference [12] models the resequencing delay caused by multipath data transfer between two hosts in a network, for which paths with distinct delays are randomly selected. Although the problem studied in [12]

is very different, the proposed formulation is applicable to the problem at hand. We refer reader to [13] and [14] for more comprehensive analysis on the subject.

Resequencing delay is defined as the time elapsed from the instant of reception of a packet by the receiver either successfully or not after the first transmission, until its disposal from the resequencing buffer. Resequencing delay, denoted by the RV $\mathbf{R}_{m,s}^i$ for the i^{th} packet can be expressed as follows.

$$\mathbf{R}_{m,s}^i = \max_{j \leq i} \{D_R \mathbf{G}_{m,s} - T_{m,s}(i-j)\}, \quad (22)$$

where $T_{m,s}$ is the mean interarrival time of packets to the receiver which are transmitted for the first time. Equation (22) formulates the resequencing delay of a packet as the maximum of either its own retransmission delay or the maximum amount of retransmission delay overlap of the preceding packets. Therefore, equation (22) has the inherent constraint $D_R Z > T_{m,s}(i-j)$, putting a limit on the number of preceding packets that may have a possible impact on the resequencing delay of the packet i . Assuming $\mathbf{R}_{m,s}^i$ to be i.i.d, defining $k = i-j$ and dropping the packet indices, (22) can be rewritten as

$$\mathbf{R}_{m,s} = \max_{0 \leq k < \frac{D_R Z}{T_{m,s}}} \{D_R \mathbf{G}_{m,s} - T_{m,s}k\}. \quad (23)$$

$T_{m,s}$ not only depends on the packet length L_m , but also on the retransmission and transmitter queue statistics as follows:

$$T_{m,s} = \frac{L_m(1 + E[\mathbf{G}_{m,s}])}{r_m(1 - V(0))}, \quad (24)$$

which is essentially the mean packet length divided by the mean capacity. Computing $X_{m,s} = \mathbb{E}[\mathbf{R}_{m,s}]$ and $\sigma_{X_{m,s}}^2 = \mathbb{V}[\mathbf{R}_{m,s}]$ based on (23) requires a few algebraic steps to derive PDF of a RV standing for the maximum of a number of statistically independent RVs with known PDFs as for (16).

III. NUMERICAL RESULTS

In this section, we evaluate the performance of the TAGLAWH scheme with a budget of Z retransmissions per packet, making its decisions irrespective of the parameters of the contending TCP flows to select the best MCS for HARQ-CC. More precisely, the proposed scheme chooses the MCS with the highest spectral efficiency having $P_{m,s,0} \leq th_P$, i.e. keeping the probability of failure in the first transmission below the threshold th_P . If no such MCS exists, TAGLAWH chooses MCS_0 . We conjecture th_P to be high enough to allow $PER_{m,s,0}$ to be calculated from online statistics without requiring any a priori channel information. We define a set of TCP traffic scenarios called $S_{N,F}$ with N flows, $N \in \{1, 2, 4, 8, 16\}$, each having a fixed RTT given by $RTT_{0,i} = 2(i+1)F/(N+1)$, where $F \in \{1, 4, 16, 64\}$ ms is the mean fixed RTT of all flows. $FER_{m,s,z}$ values are obtained through Coded Modulation Library (CML) PHY simulations based on IEEE 802.16e Wireless-MAN OFDMA PHY air interface [7]. Each MCS with Convolutional Turbo Codes (CTC) shown in Table I is simulated under the assumption of ITU Vehicular-A channel [15] from a BS to an MS (downlink) with a velocity of 90 km/hr. The resulting $FER_{m,s,z}$ curves are shown in Fig.1 for an SNR range of [0 40] dB, sampled with 2 dB resolution. Time Division Duplex (TDD) mode of WiMAX specification [16] has 35 downlink OFDM symbols

TABLE I. PARAMETERS OF MCSs SELECTED FROM IEEE 802.16.

m	V_m	R_m	$k_m(\text{bytes})$
0	4	1/2	60
1	4	3/4	54
2	16	1/2	60
3	16	3/4	54
4	64	1/2	54
5	64	2/3	48
6	64	3/4	54
7	64	5/6	60

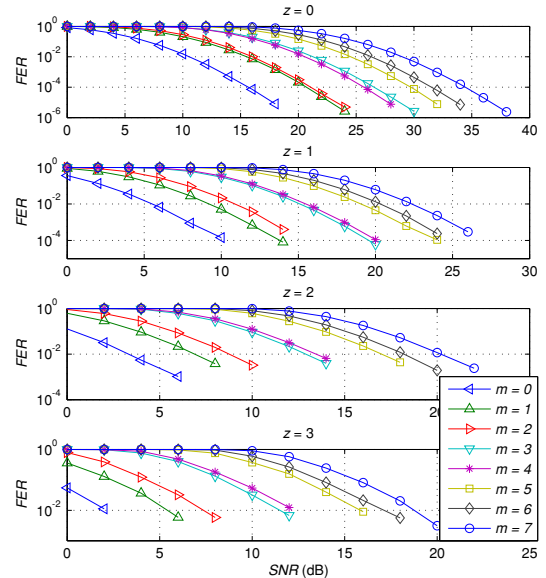


Fig. 1. Simulated FEC block error rates ($FER_{m,s,z}$) for $Z = 3$ maximum number of allowed retransmissions and the ITU Vehicular-A channel.

each consisting of 768 data sub-carriers for a TDD frame with 10 MHz channel bandwidth and 5 ms duration offering an average PHY rate of $r = 5.376 \cdot 10^6$ sub-carriers/sec. We fix the retransmission delay $D_R = 10$ ms as in [7] and the total of the two-way framing and the transmission delay $2D_F + L_m/r_m = 5$ ms.

Before evaluating the performance of TAGLAWH, we validate the proposed resequencing delay model with a number of MATLAB[®] simulations of a system consisting of: (i) a transmitter with a queue of packets to be retransmitted and having a new packet to send with a probability of $1 - V(0)$ whenever its retransmission queue is empty, (ii) a channel dropping packets as suggested by $P_{m,s,z}$ and (iii) a receiver with the aforementioned resequencing mechanism in Section II-B. We generate sixteen test cases as shown in Table II, by selecting queue occupancy atom $V(0) \in \{3 \cdot 10^{-3}, 3 \cdot 10^{-1}\}$, MCS index $m \in \{0, 2, 4, 7\}$ and SNR index s such that $P_{m,s,0} < \{3 \cdot 10^{-2}, 3 \cdot 10^{-1}\}$ is satisfied for minimum s . We present the simulation results along with the associated confidence intervals for 99% confidence levels.

There is a slight disagreement between the simulations and the proposed analysis for both the first and the second order resequencing delay statistics which can be attributed to the approximate nature of the $T_{m,s}$ expression given in (24). We also note that a packet to get retransmitted in simulation

TABLE II. RESEQUENCING DELAY VALIDATION TEST CASES AND RESULTS FOR $Z = 3$.

	$V(0)$	m	s	SNR (dB)	$P_{m,s,0}$	$X_{m,s}$ (ms)	$X_{m,s}$ -sim. (ms)	$\sigma_{X_{m,s}}$ (ms)	$\sigma_{X_{m,s}}$ -sim. (ms)
1	$3 \cdot 10^{-3}$	0	7	14	$1.26 \cdot 10^{-2}$	0.34	0.41 ± 0.04	1.55	1.80 ± 0.09
2	$3 \cdot 10^{-3}$	0	6	12	$7.84 \cdot 10^{-2}$	1.87	2.28 ± 0.08	3.28	3.77 ± 0.04
3	$3 \cdot 10^{-3}$	2	10	20	$7.66 \cdot 10^{-3}$	0.37	0.39 ± 0.06	1.58	1.62 ± 0.11
4	$3 \cdot 10^{-3}$	2	8	16	$2.25 \cdot 10^{-1}$	6.02	6.09 ± 0.05	3.53	3.57 ± 0.03
5	$3 \cdot 10^{-3}$	4	12	24	$1.51 \cdot 10^{-2}$	1.00	1.11 ± 0.10	2.46	2.63 ± 0.09
6	$3 \cdot 10^{-3}$	4	11	22	$8.79 \cdot 10^{-2}$	4.27	4.65 ± 0.09	3.74	3.92 ± 0.02
7	$3 \cdot 10^{-3}$	7	16	32	$2.34 \cdot 10^{-2}$	2.28	2.40 ± 0.12	3.32	3.43 ± 0.04
8	$3 \cdot 10^{-3}$	7	15	30	$1.15 \cdot 10^{-1}$	6.48	6.74 ± 0.13	3.13	3.17 ± 0.08
9	$3 \cdot 10^{-1}$	0	7	14	$1.26 \cdot 10^{-2}$	0.26	0.33 ± 0.02	1.38	1.65 ± 0.04
10	$3 \cdot 10^{-1}$	0	6	12	$7.84 \cdot 10^{-2}$	1.47	1.89 ± 0.05	3.08	3.62 ± 0.04
11	$3 \cdot 10^{-1}$	2	10	20	$7.66 \cdot 10^{-3}$	0.27	0.29 ± 0.03	1.38	1.43 ± 0.07
12	$3 \cdot 10^{-1}$	2	8	16	$2.25 \cdot 10^{-1}$	5.11	5.37 ± 0.07	3.87	3.90 ± 0.02
13	$3 \cdot 10^{-1}$	4	12	24	$1.51 \cdot 10^{-2}$	0.74	0.83 ± 0.05	2.18	2.36 ± 0.06
14	$3 \cdot 10^{-1}$	4	11	22	$8.79 \cdot 10^{-2}$	3.41	3.83 ± 0.12	3.73	3.99 ± 0.02
15	$3 \cdot 10^{-1}$	7	16	32	$2.34 \cdot 10^{-2}$	1.71	1.80 ± 0.08	3.03	3.15 ± 0.05
16	$3 \cdot 10^{-1}$	7	15	30	$1.15 \cdot 10^{-1}$	5.50	5.82 ± 0.10	3.54	3.63 ± 0.04

waits for an additional amount of time upon its NACK's arrival for the active transmission to finish, if any, a factor making the analytical results consistently less than those of the simulations. Overall, the accuracy of the proposed method is considered to serve well for the purposes of this paper.

Next, we evaluate performance of TAGLAWH using the presented analytical model. In Fig. 2, we show the resulting probability of failure in the first transmission $P_{m,s,0}$ of TAGLAWH w.r.t. SNR_s and th_P . TAGLAWH sticks to the most conservative MCS, MCS_0 , creating a waterfall-like region for relatively low SNR values. For increasing SNR, TAGLAWH switches to more aggressive MCSs manifesting itself as a staircase-like region until it reaches the most aggressive MCS, MCS_7 . As expected, boundary of these two regions appears at relatively lower SNR values for increasing values of th_P .

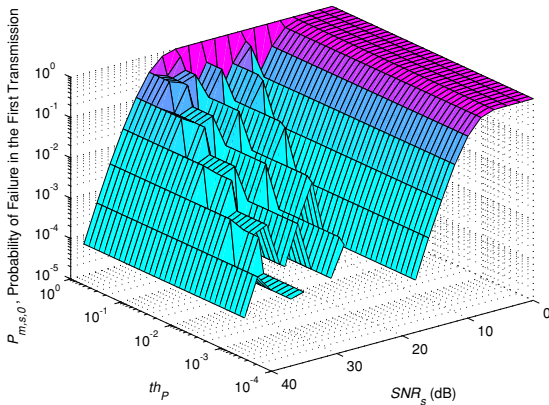


Fig. 2. Probability of failure in the first transmission $P_{m,s,0}$ of TAGLAWH for varying SNR values SNR_s and threshold parameter th_P .

We call the policy whose MCS decision gives the highest throughput for each and every value of SNR_s and traffic scenario $S_{N,F}$ as the Optimum Link Adaptation (OLA). In Fig. 3, aggregate TCP throughput of TAGLAWH normalized with that of OLA and averaged over all SNR_s values is presented for varying th_P and $S_{N,F}$ and for $Z = 3$, where traffic scenarios are indexed with the parameter idx for increasing average aggregate TCP throughput of OLA

for the sake of visualization. We discard the SNR interval corresponding to $s \in [0, s_{\min} - 1]$, such that s_{\min} is the minimum index satisfying $P_{0,s_{\min},0} < 0.5$, in all statistical calculations over SNR_s , for Fig. 3 and also for other figures to come. We find $s_{\min} = 5$ corresponding to 10 dB SNR. As th_P increases, performance of TAGLAWH reaches that of OLA regardless of the presumed traffic scenario. Owing to the retransmissions, $PER_{m,s}$ approaches zero for almost all values of th_P as evident from Fig. 1. Increasing th_P increases resequencing delay $X_{m,s}$ and so the RTT of each TCP flow. The increase in RTT, however, is outweighed by the increase in channel capacity by choosing more aggressive MCSs on the average. The reason for relatively more insensitive behavior of TAGLAWH w.r.t. th_P for low idx values is that TCP sources cannot exploit the entire PHY capacity for aggressive MCSs reaching the limit imposed by their maximum TCP window size W_{\max} .

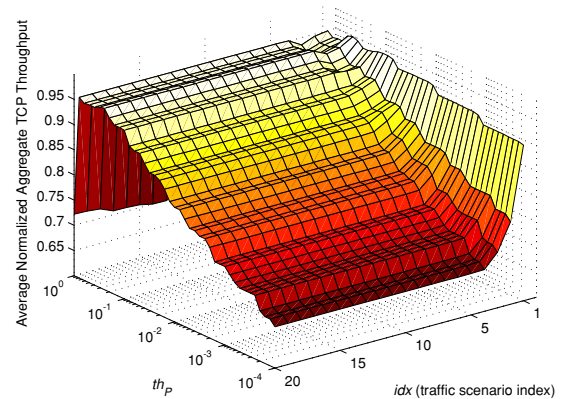


Fig. 3. Normalized aggregate TCP throughput of TAGLAWH averaged over all SNR values SNR_s for varying traffic scenarios $S_{N,F}$ and threshold parameter th_P , and for $Z = 3$ maximum number of allowed retransmissions.

In Fig. 4, both the average and the minimum (worst case) normalized aggregate TCP throughput taken over all $S_{N,F}$ and SNR_s values are given for varying maximum number of allowed retransmissions Z . TCP throughput performance remains invariant of Z regardless of the threshold th_P , since the SNR range corresponding to $s \in [0, s_{\min} - 1]$ is excluded from the calculations. For the complementary range of interest,

$s \in [s_{\min}, S - 1]$, the resulting $PER_{m,s}$ becomes negligible even after the first retransmission. In the light of the presented results, we recommend fixing th_P of TAGLAWH around 0.25, yielding an average of 4% and at worst 25% performance degradation compared to OLA. IEEE 802.16 does not enforce a particular value of th_P and this recommended value depends on the studied MCSs. Therefore, the proposed analysis should be repeated once the presumed PHY technology is changed.

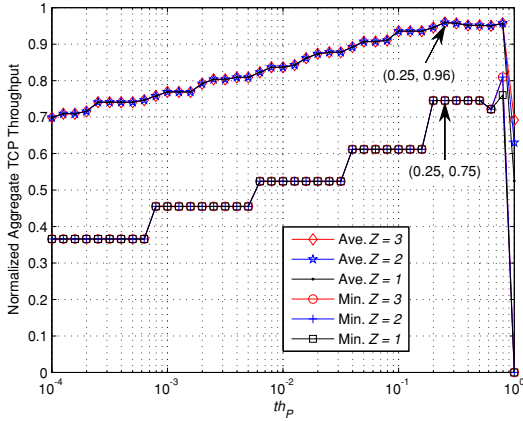


Fig. 4. Average and minimum (worst case) normalized aggregate TCP throughput of TAGLAWH taken over all SNR values SNR_s and traffic scenarios $S_{N,F}$ for varying threshold parameter th_P and maximum number of allowed retransmissions Z .

Finally, both the average and the maximum (worst case) values of mean resequencing delay $X_{m,s}$ are shown in Fig. 5 again for varying th_P and Z . For the same arguments made for the results of Fig. 4, $X_{m,s}$ remains unaffected for the choice of Z until th_P approaches 1, for which the minimum normalized aggregate TCP throughput drops down to 0. For the proposed value of $th_P = 0.25$, the maximum (worst case) and the average values of $X_{m,s}$ are computed to be around 6.49 ms and 2.50 ms, respectively. As long as the wireless link of interest remains the bottleneck link for all TCP flows it is serving for, th_P parameter of TAGLAWH can be optimized for throughput without paying much attention to increasing RTT due to the resequencing delay. For the estimation of the resulting throughput at a particular instance of the channel condition and the traffic scenario, however, resequencing delay needs to be taken into account.

IV. CONCLUSIONS

In this paper, we propose an analytical method to evaluate the TCP level throughput performance of a wireless bottleneck link deploying AMC with HARQ. Based on the assumptions and results presented in this paper, we propose to maintain the probability of failure in the first transmission attempt at a value of 0.25 when using IEEE 802.16 MCSs. Future work remains in terms of analysis of TCP throughput in a network of wireless links relying on HARQ where the bottleneck link may change depending on the parameters of each link.

REFERENCES

[1] M. Laor and L. Gendel, "The effect of packet reordering in a backbone link on application throughput," *Network, IEEE*, vol. 16, no. 5, pp. 28–36, Sep 2002.

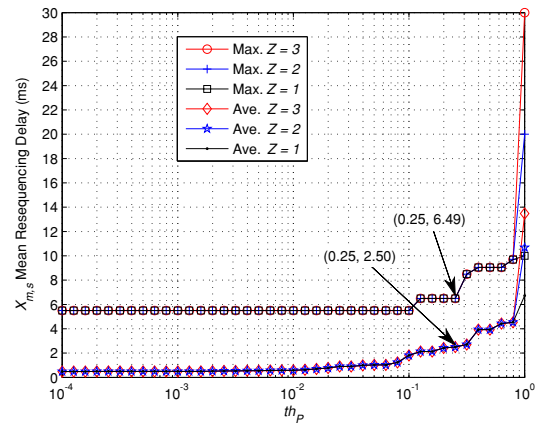


Fig. 5. Average and maximum (worst case) mean resequencing delay $X_{m,s}$ of TAGLAWH taken over all SNR values SNR_s and traffic scenarios $S_{N,F}$ for varying threshold parameter th_P and maximum number of allowed retransmissions Z .

[2] J. Ramis and G. Femenias, "Cross-layer design of adaptive multirate wireless networks using truncated HARQ," *Vehicular Technology, IEEE Transactions on*, vol. 60, no. 3, pp. 944–954, Mar 2011.

[3] D. Moltchanov, "A study of TCP performance in wireless environment using fixed-point approximation," *Computer Networks*, vol. 56, no. 4, pp. 1263–1285, Mar 2012.

[4] K.-C. Go, J.-R. Cha, S. K. Oh, and J.-H. Kim, "End-to-end performance analysis based on cross-layer retransmission scheme in wireless communication system," *The International Conference on Information Networking 2013 (ICOIN)*, pp. 141–144, Jan 2013.

[5] O. Ozturk and N. Akar, "Workload-dependent queuing model of an AQM-controlled wireless router with TCP traffic and its application to PER-based link adaptation," *EURASIP Journal on Wireless Communications and Networking*, vol. 67, no. 1, Apr 2014.

[6] R. Bekker, S. Borst, O. Boxma, and O. Kella, "Queues with workload-dependent arrival and service rates," *Queueing Systems*, vol. 46, no. 3-4, pp. 537–556, Mar 2004.

[7] "IEEE standard for local and metropolitan area networks part 16: Air interface for broadband wireless access systems," *IEEE Std 802.16-2009 (Revision of IEEE Std 802.16-2004)*, pp. 1–2080, 2009.

[8] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose, "Modeling TCP throughput: A simple model and its empirical validation," *SIGCOMM Comput. Commun. Rev.*, vol. 28, no. 4, pp. 303–314, Oct 1998.

[9] (2014) ns-3. <http://www.nsnam.org/>.

[10] V. Vasudevan, A. Phanishayee, H. Shah, E. Krevat, D. G. Andersen, G. R. Ganger, G. A. Gibson, and B. Mueller, "Safe and effective fine-grained TCP retransmissions for datacenter communication," *SIGCOMM Comput. Commun. Rev.*, vol. 39, no. 4, pp. 303–314, Aug 2009.

[11] S. Floyd, "Recommendation on using the Gentle variant of RED," Tech. Rep., Mar 2000.

[12] K. Zheng, X. Jiao, M. Liu, and Z. Li, "An analysis of resequencing delay of reliable transmission protocols over multipath," in *Communications (ICC), 2010 IEEE International Conference on*, May 2010, pp. 1–5.

[13] J. Li and Y. Zhao, "Resequencing analysis of Stop-and-Wait ARQ for parallel multichannel communications," *Networking, IEEE/ACM Transactions on*, vol. 17, no. 3, pp. 817–830, Jun 2009.

[14] N. Shacham and B.-C. Shin, "A selective-repeat-ARQ protocol for parallel channels and its resequencing analysis," *Communications, IEEE Transactions on*, vol. 40, no. 4, pp. 773–782, Apr 1992.

[15] "Guidelines for evaluation of radio transmission technologies for IMT-2000," *Recommendation ITU-R M.1225*, 1997.

[16] "WiMAX Forum mobile system profile: Release 1.5 TDD specific part," Tech. Rep., Mar 2009.