

Cross-layer queueing analysis for aggregated ON-OFF arrivals with adaptive modulation and coding

Xingyu Han, John Schormans
 Department of Electronic Engineering and Computer Science
 Queen Mary, University of London, E1 4PH, UK
 x.han@qmul.ac.uk, j.schormans@qmul.ac.uk

Abstract: Packet loss and delay in wireless networks is a function of both the traffic characteristics (e.g. load, burstiness) and the characteristic behavior of the fading channel. In this paper we propose and analyze a new queue model in which the server process is controlled by the adaptive modulation and coding (AMC) algorithm, and with aggregated ON-OFF packet arrivals to a finite buffer. In the scenarios where users require multiple services, aggregated ON-OFF arrivals approximate bursty behavior better than the Poisson arrivals analyzed in many earlier queueing models. We focus on the packet drop probability (PDP), the packet loss probability (PLP), the average queueing delay and the throughput, showing significant differences in these metrics between the cross-layer models using aggregated ON-OFF arrivals and Poisson arrivals. This indicates that in wireless packet network performance evaluation, the traffic model is at least as important as the effect of the fading channel modelling.

1. INTRODUCTION

Multimedia packet-based wireless networks require high data rates for the provision of higher spectral efficiency and lower packet loss probability. Because of the properties of wireless channels, the impact of fading must be considered. Most of the existing networks use adaptive modulation and coding (AMC) at the physical layer as a means of combating channel quality variation, i.e. as the channel worsens the modulation and coding scheme (MCS) transmits fewer bits per symbol to maintain the target BER.

However, traditional AMC operating at the physical layer doesn't consider the impact of buffering at the data-link layer, so earlier analyses do not properly approximate the system behaviour. Thus, queueing analysis which takes into account the conditions of the data-link layer becomes necessary. Recently, cross-layer analysis combining queueing effects at the data-link layer and AMC at the physical layer have received a lot of attention in various scenarios [3]-[8], [26]-[34], including single-user (SU) and multi-user (MU) scenarios with MIMO or multi-hop relay. Combining the data-link and the physical layer mechanisms makes it possible to analyze QoS provisioned metrics such as packet loss, packet delay and throughput.

At the transmitter side, the above mentioned cross-layer analysis can be well modelled with queueing theory. In fact, most of the cross-layer analysis found in published works [3]-[8], [26]-[34] use working-vacation queue models. Typical queue models with working vacations [1] assume a lower service rate during vacations compared to the non-vacation period, instead of the server completely stopping as in a conventional

vacation model. The published cross-layer analysis adopts a range of working vacations with different service rates, such as to fit with the MCSs used at the physical layer. More specifically, most of the analyses of cross-layer designs found in the existing literature [3]-[8] utilize an M/D/s(t)/K queue model, in which the buffer service rate, which is directly controlled by the MCSs, varies over time and in which arrivals follow a Poisson process.

However, for users requiring multiple streams, which means multiple bursty packet streams arriving and getting served simultaneously, the Poisson model is no longer accurate. Although publications such as [33] and [34] introduce discrete-time batch Markovian arrival process (DBMAP) as an arrival traffic model to replace the Poisson process, they didn't investigate the bursty behavior of such arrival models. Therefore, analyzing the arrival processes which can better approximate the practical communication system with multiple bursty services still remains as an open question in cross-layer analysis.

There have been extensive investigations on traffic models in the existing literature [14]-[15], [17], [19]-[24]. Most of them suggest that the ON-OFF traffic model is one of the most appropriate models to approximate bursty packet streams. The single ON-OFF traffic model, which gives a higher constant arrival rate during on periods and a lower constant arrival rate during off periods, is naturally suitable for bursty voice streams [15], [17], [24]. In addition, the superposition of multiple single ON-OFF traffic model, which is also known as aggregated ON-OFF model (or N-burst model, Batch ON-OFF model etc.), is good at approximating bursty data traffics based on extensive observations in existing literature [14], [18]-[23]. In fact, the single ON-OFF traffic model is a special case of the aggregated ON-OFF one. All of these published papers give strong support to use the aggregated ON-OFF traffic model instead of the Poisson arrival model to approximate bursty streams, making aggregated ON-OFF arrival model one of the most appropriate choices when applied to queueing analysis on bursty streams.

In this paper, we analyze the queueing behavior of the cross-layer analysis combining aggregated ON-OFF arrival at the data-link layer and AMC at the physical layer. Then we construct a finite-state Markov chain (FSMC) with a state tuple including buffer state, arrival state and service state, and show how to get the stationary distribution of the probability of the queue length in the buffer using matrix analytic method. Then, we validate the queue state through Monte-Carlo simulation, laying the foundation from which we obtain metrics including the PDP due to buffer overflow, the average queueing delay, the PLP and the throughput. Finally, we illustrate the impact of bursty traffic on the proposed cross-layer analysis.

The rest of the paper is constructed as follows. Section 2 presents the system model for the cross-layer analysis. Performance metrics and feasibility discussions for the proposed cross-layer analysis are given in Section 3. Numerical results are obtained in Section 4. And in Section 5, we discuss the possible extensions to the proposed cross-layer analysis. Section 6 draws conclusions.

2. SYSTEM MODEL

2.1 System Overview

An end-to-end wireless link with multiple antennas at both transmitter and receiver side is considered. At the transmitter side, arriving packets at the data-link layer feed a queue with finite buffer size K . The packets are converted into frames and are served in a first-in-first-out (FIFO) manner. The frame duration T_f (also referred to as a timeslot) is fixed, and the number of packets per frame depends on the MCS selected. With the introduction of an OFDM block at the physical layer, which helps combat the frequency selective fading, we assume a slow and frequency flat Rayleigh fading channel, so that the mode used during one frame duration remains the same. The MCS changes at the end of each frame, in accordance with a Markov process. At the receiver side, a channel estimator measures the current Channel State Information (CSI), and an MCS selector at the receiver side then determines the MCS used during next frame duration and feeds it back to the transmitter side through an error-free feedback channel.

At the air interface, we consider a point to point SU-MIMO channel with M_t transmit antennas and M_r receive antennas. The channel is modeled as,

$$\mathbf{y} = \sqrt{E_S} \mathbf{H} \mathbf{x} + \mathbf{n} \quad (1)$$

where E_S is the transmit energy per antenna; \mathbf{y} is $M_r \times 1$ receive vector; \mathbf{x} is $M_t \times 1$ transmit vector with each element x_k selected from unit energy constellation; \mathbf{n} is $M_r \times 1$ circular symmetric complex additive white Gaussian noise vector with variance $N_0 \mathbf{I}_{M_r}$ where \mathbf{I}_{M_r} is the identity matrix; \mathbf{H} is the $M_r \times M_t$ channel gain matrix with each element h_{ij} the channel gain from the j th transmit antenna and i th receive antenna.

We adopt the zero-forcing detector at the receiver side. The received $M_r \times 1$ symbol vector $\hat{\mathbf{y}}$ after pseudo-inverse operation is shown by,

$$\hat{\mathbf{y}} = \mathbf{G} \mathbf{y} = \mathbf{G} (\sqrt{E_S} \mathbf{H} \mathbf{x} + \mathbf{n}) = \mathbf{x} + \mathbf{G} \mathbf{n} \quad (2)$$

where $\mathbf{G} = \frac{1}{\sqrt{E_S}} \mathbf{H}^\dagger = \frac{1}{\sqrt{E_S}} (\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H$ denotes the pseudo-inverse operation to channel gain matrix \mathbf{H} so that $\sqrt{E_S} \mathbf{G} \mathbf{H} = \mathbf{I}$; \mathbf{H}^H denotes the Hermitian transpose or conjugate transpose of channel gain matrix \mathbf{H} .

Therefore, the SNR on the k th spatial stream γ_k can be calculated by,

$$\gamma_k = \frac{p(x_k x_k^H)}{p([\mathbf{G} \mathbf{n}][\mathbf{G} \mathbf{n}]^H]_{kk}} = \frac{E_S}{N_0 [\mathbf{H}^\dagger \mathbf{H}^\dagger]_{kk}} = \frac{E_S}{N_0 [(\mathbf{H}^H \mathbf{H})^{-1}]_{kk}} = \frac{\gamma_0}{[(\mathbf{H}^H \mathbf{H})^{-1}]_{kk}} \quad (3)$$

where $\gamma_0 = \frac{E_S}{N_0}$; $p(\cdot)$ is the symbol or noise power; $[\cdot]_{kk}$ is the k th diagonal entry of the matrix. Note that $p(x_k x_k^H) = 1$ since each symbol x_k is picked from a unit energy constellation.

Next, we adopt the Kronecker model as the spatial correlation model. Generally, the channel gain matrix \mathbf{H} for the Kronecker model is obtained by,

$$\mathbf{H} = \mathbf{R}_r^{1/2} \mathbf{H}_w (\mathbf{R}_t^{1/2})^\top \quad (4)$$

where the elements of \mathbf{H}_w are independent and identical complex Gaussian random numbers, distributed with zero mean and unit variance; \mathbf{R}_t is the transmit correlation matrix; \mathbf{R}_r is the receive correlation matrix; \top is the transpose of the matrix.

If we only consider transmit correlation assuming there is rich scattering at the receiver side, since each row of the channel gain matrix \mathbf{H} follows an M_t -variate Normal distribution with zero mean, $(\mathbf{H}^H \mathbf{H})$ is a complex Wishart matrix. Therefore, as suggested by [35], the SNR on the k th spatial stream γ_k with zero-forcing detector can be obtained by,

$$f(\gamma_k) = \frac{[\mathbf{R}_t^{-1}]_{kk} e^{-\frac{\gamma_k [\mathbf{R}_t^{-1}]_{kk}}{\gamma_0}}}{\gamma_0 \Gamma(M_r - M_t + 1)} \left(\frac{\gamma_k [\mathbf{R}_t^{-1}]_{kk}}{\gamma_0} \right)^{M_r - M_t} \quad (5)$$

However, for the scenario with both transmit and receive correlations, each row of the channel gain matrix \mathbf{H} given by (4) does not follow a multi-variate Normal distribution. Therefore, it is at least very hard, even if possible, to obtain the closed-form expressions for the distribution of the SNR on the k th spatial stream γ_k . Therefore, we evaluate γ_k by using Monte-Carlo simulation instead. More specifically, we firstly capture the CDF of the SNR for each stream. Then we fit these curves using exponential functions. After derivations of these fitted exponential functions, we obtain the asymptotic expressions for the distribution of the SNR for each stream.

2.2 Adaptive Modulation and Coding

The AMC algorithm adopts a range of MCSs, also referred to as modes, to adaptively adjust to the current CSI. As previously mentioned, we assume the fading channel to be frequency flat because of OFDM operation, so that the channel condition is unchanged during each frame, but may vary between adjacent frames.

AMC tries to maximize the data rate given a prescribed system PER P_0 . PER is caused by the noise during transmission. We assume an additive Gaussian white noise (AWGN) channel in our system. More specifically, AMC determines the signal-to-noise ratio (SNR) thresholds for each mode to transmit. The adopted modes are listed in TABLE I.

Parameters a_n, g_n, γ_{pn} , which depend on parameter settings for PER analysis listed in TABLE II, are related to asymptotic expression for PER which are specified in [4] given by,

$$PER_n(\gamma) = \begin{cases} a_n \times e^{-g_n \times \gamma}, & \text{for } \gamma \geq \gamma_{pn} \\ 1, & \text{otherwise} \end{cases} \quad (6)$$

where γ is SNR and n denotes the transmission mode. Then the threshold γ_n for mode n is given by,

$$\gamma_n = \frac{-1}{g_n} \times \ln\left(\frac{P_0}{a_n}\right) \quad (7)$$

TABLE I
Modes Adopted

Mode n	$n=0$	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$
MCS Index	-	0	2	4	6	8
Modulation	-	BPSK	QPSK	16QAM	64QAM	256QAM
Coding rate	-	1/2	3/4	3/4	3/4	3/4
Coded bits per symbol	0	1	2	4	6	8
Data bits per symbol	0	0.5	1.5	3	4.5	6
Service rate (packets/slot)	0	1	3	6	9	12
a_n	0	2.898	2.690	2.973	2.934	2.999
g_n	0	0.7383	0.2041	0.04633	0.01158	0.003091
$\gamma_{pn}(dB)$	-	1.5872	6.8559	13.7139	19.6825	25.6085

Note that the system may also use a mode 0 with zero data rate, which means the instantaneous channel condition is too poor for transmission. As noted earlier, the pdf for the fading channel follows (5) so that the probability to select mode n for the k th stream is denoted as,

$$\pi_k^n = \int_{\gamma_n}^{\gamma_{n+1}} f(\gamma_k) d\gamma_k \quad (8)$$

Then, we can approximate the average PER for the k th stream \overline{PER}_k which is shown by,

$$\overline{PER}_k = \frac{\sum_{n=1}^N R_n \times \pi_k^n \times \overline{PER}_k^n}{\sum_{n=1}^N R_n \times \pi_k^n} \quad (9)$$

where R_n is the data bits per symbol given by TABLE I, and \overline{PER}_k^n is the average PER in mode n for the k th stream. We can make sure the system PER is less than the prescribed P_0 by setting each $\overline{PER}_k^n \leq P_0$ and obtaining the thresholds γ_n .

Since we assume a slow fading channel, the mode transition between successive timeslots can be formulated as a Markov chain of quasi-birth-death process; accordingly the transition probability matrix for mode transition with each element $p_{m,n}^k$ (probability to transfer from mode m to mode n) can be generated as,

$$\begin{cases} p_{n,n+1}^k = \frac{N_{n+1} \times T_f}{\pi_k^n}, \text{ if } n \in [0, N-1], n \in Z \\ p_{n,n-1}^k = \frac{N_n \times T_f}{\pi_k^n}, \text{ if } n \in [1, N], n \in Z \end{cases} \quad (10)$$

where N_n is level crossing rate (LCR) which is evaluated by Monte-Carlo simulation, and T_f is calculated by,

$$T_f = T_s \times \text{ceil} \left(\frac{\text{Symbols per frame}}{\text{Number of data subcarriers}} \right) \quad (11)$$

where $T_s = \text{OFDM symbol duration} + \text{GI duration} = 4 \mu\text{s}$, $\text{ceil}(\cdot)$ is the ceiling function. Other parameters can be obtained in TABLE II.

TABLE II
Parameter Settings for PER Analysis

Parameter	Value	Parameter	Value
Packet length	1080 bits	Symbols per frame	2160
SNR range	-5~35 dB	Channel bandwidth	20 MHz
No. of data subcarriers	52	No. of FFT points	64
Bits/symbol	[1,2,4,6,8]	Interleaving block	$(2 \times \text{bits/sym}) \times 26$
Numerical system	Grey code	Modulation	Rectangular QAM
OFDM symbol duration	$3.2 \mu\text{s}$	GI	800 ns

3. PERFORMANCE ANALYSIS

In this section, we focus on the queueing analysis of the proposed N-ON-OFF/D/s(t)/K queue model, in which the system accepts multiple ON-OFF arriving packets, and serves the incoming packets with a varied number of servers determined by the AMC and the fading channel, and has a finite buffer length of K packets. Arrival process, service process and queue state transitions are discussed for the proposed queue model, and finally we construct a finite-state Markov chain (FSMC) for queue state transition probabilities. Moreover, we address the feasibility of obtaining the stationary distribution of queue state based on the queue state transition matrix.

At the transmitter side, we assume an independent buffer for each antenna. Therefore, the performance analysis for each stream is independent from each other. For convenience, we remove the sub-index k for the k th stream for performance analysis.

The whole system is slot-based. We denote timeslot as t , each of which has frame duration of T_f . At most S_t packets (as determined by the MCS selected) are transmitted at the beginning of timeslot t . During timeslot t , A_t packets arrive into the system. We also denote by B_t the queue state at timeslot t , this is calculated after packet departures and arrivals.

3.1 Arrival process

The adopted arrival process A_t of the proposed cross-layer analysis is aggregated ON-OFF process, which is independent of the queue state and the service process. As shown in Fig. 1, the buffer can accept at

most P_I streams, each of which P_s follows an ON-OFF process with c_{P_s} packets/slot when it's in the ON state and 0 packets/slot in the OFF state. Each of these ON-OFF processes follows an independent Markov chain (MC) with probability transition matrix,

$$\mathbf{P}_{on-off}^{P_s} = \begin{pmatrix} p_0^{P_s} & 1 - p_0^{P_s} \\ 1 - p_1^{P_s} & p_1^{P_s} \end{pmatrix} \quad (12)$$

where state 0 is OFF, and state 1 is ON. $p_0^{P_s}$ is the probability for the process to stay in the OFF state, and $p_1^{P_s}$ is the probability to stay in the ON state. Therefore, we can obtain the stationary distribution for the ON and OFF states of each ON-OFF process by solving,

$$\boldsymbol{\pi}_{P_s} = \boldsymbol{\pi}_{P_s} \times \mathbf{P}_{on-off}^{P_s}, \quad \sum_{P_s \in [1, P_I]} \boldsymbol{\pi}_{P_s} = 1 \quad (13)$$

We obtain $\pi_{P_s}^{on}$ and $\pi_{P_s}^{off}$ as stationary probability for stream P_s in ON state and OFF state.

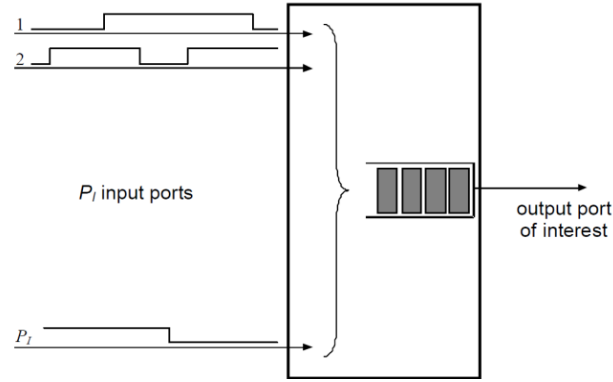


Fig. 1 Aggregated ON-OFF arrival model [9]

All of the single ON-OFF arrivals of these P_I streams will combine together, and the outcome outputs an aggregated ON-OFF arrival process. The average arrival rate for aggregated ON-OFF arrival process \bar{A}_t can be calculated as,

$$\bar{A}_t = \sum_{P_s=1}^{P_I} \pi_{P_s}^{on} \times c_{P_s} \quad (14)$$

Next, we calculate the probability of k single streams being in the ON or OFF state N^{on} and N^{off} by analyzing each individual $\mathbf{P}_{on-off}^{P_s}$. Assuming $\{\mathbf{T}_k\}$ is the set of all the combination of k streams in ON state with $\binom{P_I}{k}$ elements, and denote each combination as \mathbf{T}_k^i , and each complementary combination as $\mathbf{C}(\mathbf{T}_k^i)$, then,

$$\left\{ \begin{array}{l} \Pr(N^{on} = k) = \sum_{i=1}^{\binom{P_I}{k}} \left(\prod_{m \in T_k^i} \pi_m^{on} \times \prod_{n \in C(T_k^i)} \pi_n^{off} \right) \\ \Pr(N^{off} = k) = \sum_{i=1}^{\binom{P_I}{k}} \left(\prod_{m \in T_{P_I-k}^i} \pi_m^{on} \times \prod_{n \in C(T_{P_I-k}^i)} \pi_n^{off} \right) \end{array} \right. \quad (15)$$

Note that the aggregated ON-OFF arrival model we adopt in this sub-section is readily applied to multi-user scenarios such as WLAN, where one access point (AP) with one finite buffer can support multiple users simultaneously. Regardless of the destinations of the incoming packets at the AP, the arrival process can be well modelled by the aggregated ON-OFF process. And we can obtain the same performance analysis which is detailed in the rest of section 3 for both single-user and multi-user scenarios. However, since we only consider single-user scenario, we use the single ON-OFF as the arrival model for section 4. For convenience, we denote both single and aggregated ON-OFF arrival models as aggregated ON-OFF arrival model because aggregated ON-OFF also includes single ON-OFF arrival model.

3.2 Service process

Service process S_t is determined by the AMC algorithm operating at the physical layer in order to adjust to CSI. The packets are reassembled into frames for transmission. Therefore the number of packets that gets transmitted per frame is totally determined by the mode selected, which is equivalent to a deterministic service process with varied number of servers. The number of servers can only be chosen from a set \mathcal{S} , of which,

$$S_t \in \mathcal{S}, \mathcal{S} = \{s_n, n \in [0, N]\}, s_n = b \times R_n \quad (16)$$

Normally, we keep $\{s_n\}$ as integer values, thus, b is usually set to 2. As a result, $\mathcal{S} = \{0, 1, 3, 6, 9, 12\}$.

3.3 Queue State Transition

As mentioned above, B_t is the queue state at the end of a timeslot t , or equivalently, the queue state at the beginning of timeslot $t+1$. At the beginning of timeslot t , the system serves at most S_t packets, or serves B_{t-1} packets if $B_{t-1} < S_t$. Then the aggregated ON-OFF arrival packets at timeslot t A_t enters the queue. If the current queue state after arrivals is larger than the buffer size K , then $B_t = K$ and excess packets will be dropped due to buffer overflow. The queue state transition process can be calculated by,

$$B_t = \min(K, \max(0, (B_{t-1} - S_t)) + A_t) \quad (17)$$

Then we can obtain the probability transition matrix for queue state \mathbf{P}_B as follows,

$$\mathbf{P}_B = \{p_{u,v}\} = \{p(B_t = v | B_{t-1} = u)\}, 0 \leq u, v \leq K \quad (18)$$

More specifically, we denote by A , which is a vector, all possible summations of combinations of single ON-OFF streams $\{c_{P_s}\}$, and $A_t \in A$. Since the next states for both of the arrival and service processes are dependent on the current states, we have to consider all possible single cases for both of arrival and service process which is shown by,

$$p_{u,v} = \sum_{A_t, A_{t+1} \in A, S_t, S_{t+1} \in \mathcal{S}} p(A_{t+1} = b, B_t = v, S_{t+1} = d | A_t = a, B_{t-1} = u, S_t = c) \quad (19)$$

For each $p_{(a,u,c),(b,v,d)}$,

$$p_{(a,u,c),(b,v,d)} = p(A_{t+1} = b | A_t = a) \times p(S_{t+1} = d | S_t = c) \times p(B_t = v | A_t = a, B_{t-1} = u, S_t = c) \quad (20)$$

$p(A_{t+1} = b | A_t = a)$ is the transition probability for the aggregated ON-OFF packet arrivals. If all of $c_{P_s} = 1$, then its calculation can be obtained from (15) as follows,

$$p(A_{t+1} = b | A_t = a) = \Pr(N^{on} = a) \times \Pr(N^{on} = b) \quad (21)$$

(21) is obtained because the arrival rates are assumed to be independent between adjacent timeslots, and the arrival rate equals to the number of ON-state streams when $c_{P_s} = 1$.

$p(S_{t+1} = d | S_t = c)$ is the transition probability for the service rate process, and $p(B_t = v | A_t = a, B_{t-1} = u, S_t = c)$ is obtained as follows,

$$p(B_t = v | A_t = a, B_{t-1} = u, S_t = c) = \begin{cases} 1, & \text{if } v = \min(K, \max(0, (u - c)) + a) \\ 0, & \text{if } v \neq \min(K, \max(0, (u - c)) + a) \end{cases} \quad (22)$$

Next, we have to discuss the feasibility of getting the stationary distribution of the queue state. As shown in [3], the stationary distribution $\boldsymbol{\pi}$ exists and is unique if the probability transition matrix of the enlarged FSMC is irreducible, homogeneous and positive recurrent. Then we can have the lemma below,

Lemma: The Markov chain of (a, u, c) , where $(a, u, c) \in A \times B \times S$, has only one closed communicating class, and therefore is positive recurrent.

Proof: Firstly, we need to show there exists a multi-transition path with non-zero transition probability from state $(a, u, c) \rightarrow (b, v, d)$, where $(a, u, c), (b, v, d) \in A \times B \times S$.

1) If $v = \min(K, \max(0, (u - c)) + a)$, we can find a direct path from $(a, u, c) \rightarrow (b, v, d)$ with non-zero transition probability by (5.19).

2) If $v \neq \min(K, \max(0, (u - c)) + a)$, there exists a path with non-zero transition probability from $(a, u, c) \rightarrow (b', v', d')$, where either b' or d' is equal to 0; then we can always find a multi-transition path from $(b', v', d') \rightarrow (b, v, d)$ with each intermediate state (b'', v'', d'') , where either b'' or d'' is equal to

0, to offset the difference between v' and v .

Then we can draw the conclusion that the finite state set $(a, u, c) \in \mathbf{A} \times \mathbf{B} \times \mathbf{S}$ forms a closed communicating class where every pair of states $(a, u, c), (b, v, d)$ in the set communicates with each other. Therefore, the state transition matrix is irreducible by definition. In addition, since the FSMC of state transition is independent of time, the state transition matrix is homogeneous by definition.

Finally, we can conclude that the Markov chain of state transition is positive recurrent, because [3] asserts that the finite state irreducible homogeneous Markov chain is positive recurrent. (*End of Proof*)

Then we can draw the conclusion that the stationary distribution of queue state $\boldsymbol{\pi}_{P_B}$ exists and is unique, and $\boldsymbol{\pi}_{P_B} \geq 0$.

Based on (18)-(22), we can get the stationary distribution of queue state by solving,

$$\boldsymbol{\pi}_{P_B} = \boldsymbol{\pi}_{P_B} \times \mathbf{P}_B, \quad \sum_{a \in \mathbf{A}} \sum_{c \in \mathbf{S}} \sum_{u \in [0, K]} \boldsymbol{\pi}_{P_B} = 1 \quad (23)$$

3.4 Performance Analysis

We would like to analyze the PDP due to buffer overflow, the average queueing delay, the PLP (caused by PDP and transmission PER) and the system throughput for the proposed cross-layer analysis. Firstly, we need to obtain the packet drop process, denoted D_t . As mentioned above, the excess packets are dropped within each timeslot 't' if the buffer is full, thus, we can obtain the expression for D_t as follows,

$$D_t = \max(0, A_t - K + \max(0, B_{t-1} - S_t)) \quad (24)$$

Thus, the average number of dropped packets during timeslot t is,

$$E(D) = \sum_{A_{t+1} \in \mathbf{A}} \sum_{B_{t-1}=0}^K \sum_{S_t=S_0}^{s_n} D_t \times P(A_t = a, B_{t-1} = u, S_t = c) \quad (25)$$

Next we get the PDP due to buffer overflow as follows,

$$PDP = \frac{E(D)}{\bar{A}_t} = \frac{\sum_{A_{t+1} \in \mathbf{A}} \sum_{B_{t-1}=0}^K \sum_{S_t=S_0}^{s_n} D_t \times P(A_t = a, B_{t-1} = u, S_t = c)}{\sum_{P_s=1}^{P_I} \pi_{P_s}^{on} \times c_{P_s}} \quad (26)$$

Next, we can obtain the average queueing delay \bar{W} based on Little's Law [25]. The expression for \bar{W} is shown as follows,

$$\bar{W} = \frac{E(B)}{(1 - PDP) \times \bar{A}_t} \quad (27)$$

where $E(B)$ is the average queue length which can be obtained by,

$$E(B) = \sum_{B=0}^K \pi_{P_B} \times B \quad (28)$$

Then, we can evaluate the PLP as follows,

$$PLP = 1 - (1 - PDP) \times (1 - \overline{PER}) \quad (29)$$

And finally, we can obtain the system throughput by,

$$T = \overline{A}_t \times (1 - PLP) \quad (30)$$

4. NUMERICAL RESULTS

In this section, we present numerical results based on analytical expressions from Section 2 and 3. We concentrate on making comparisons between Poisson arrivals and aggregated ON-OFF arrivals for performance analysis.

We assume the prescribed PER $P_0 = 10^{-2}$ (note that P_0 is the upper bound, and average PER \overline{PER} is much less than P_0 , approximately equal to 3.4587×10^{-6} and 5.3616×10^{-6} for Poisson and aggregated ON-OFF with the MCSs adopted in TABLE I), buffer lengths $K=100$ for queue state analysis and $K=200$ for other performance analysis as suggested by [36].

The probability transition matrix $P_{on-off}^{P_s}$ is chosen to approximate the burstiness of each single ON-OFF stream. As suggested in [13], a practical packet generation model for wireless networks is given by a DBMAP for ON-OFF traffic, which should follow the probability transition matrix as follows,

$$P_{on-off}^{P_s} = \begin{pmatrix} 0.9892 & 0.0108 \\ 0.0143 & 0.9857 \end{pmatrix} \quad (31)$$

Note that the values provided by (31) correspond to individual subscriber Internet scenario as suggested by [13], and these values are captured from real network studies in order to obtain the distributions associated with HTTP sessions. Although the experiments carried out in [13] were not intended to be applied at the data-link layer, they provide the basic idea of bursty traffic models, which captures the burstiness of network traffic at all levels.

The matrix (26) suggests that the probabilities for single ON-OFF arrival staying in the on and off state are so high (0.9857 and 0.9892 respectively) that the state rarely changes, which is a reasonable approximation for bursty behavior, since arriving packets will keep coming during a long period of time while staying in the on state, and then no packets will arrive during another long period of time staying in the off state.

In addition, we assume the arrival rate during ON periods of the single ON-OFF model c_{P_s} is an integer

value. Therefore, we can obtain the single ON-OFF stream arrival rate as a multiple of $\pi_{P_S}^{on}$.

Finally, the proper transmit and receive correlation matrices \mathbf{R}_t and \mathbf{R}_r need to be adopted. We check a 4×4 SU-MIMO scenario for performance analysis. Consistent with those adopted in [35], we set \mathbf{R}_t and \mathbf{R}_r as,

$$\mathbf{R}_t = \mathbf{R}_r = \begin{pmatrix} 1 & 0.57e^{-2.25i} & 0.17e^{0.02i} & 0.29e^{-2.94i} \\ 0.57e^{2.25i} & 1 & 0.57e^{-2.25i} & 0.17e^{0.02i} \\ 0.17e^{-0.02i} & 0.57e^{2.25i} & 1 & 0.57e^{-2.25i} \\ 0.29e^{2.94i} & 0.17e^{-0.02i} & 0.57e^{2.25i} & 1 \end{pmatrix}, \quad (32)$$

Note that the correlation matrices are not necessarily the same, and the elements for correlation matrices are also dependent on real antenna configurations, which may vary in different scenarios.

Fig. 2 shows the CDF of the SNR for each stream when we adopt (32). We obtain that the curves for stream 1 & 4 are the same; while the same situation goes to stream 2 & 3. With the same arrival rate and channel model, we obtain the performance metrics are mainly dependent on the service rate, which is determined by the CDF of the SNR. Therefore, we only need to check stream 1 and stream 2 for simplicity.

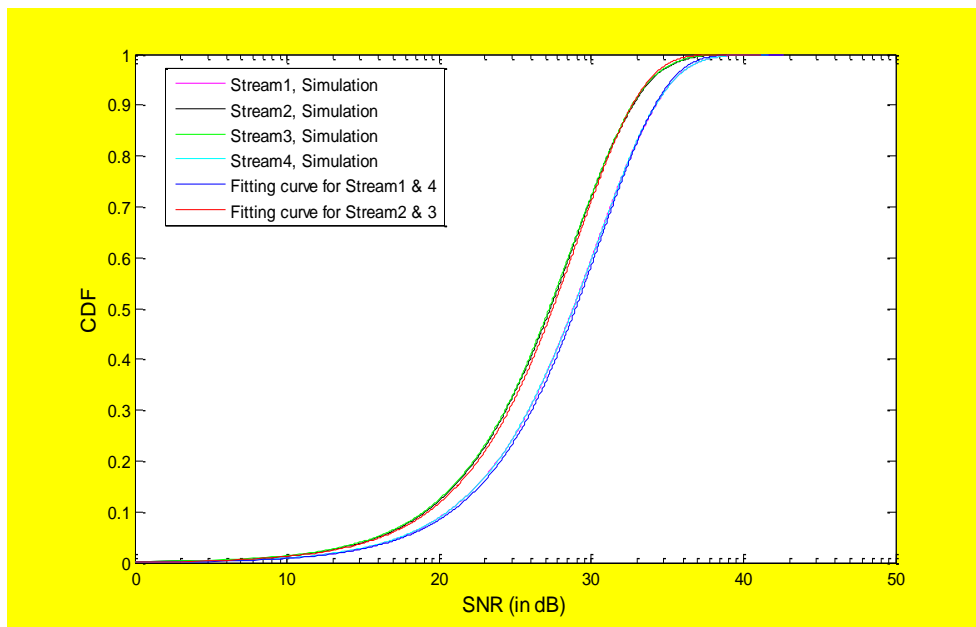


Fig. 2 CDF of the SNR for each stream

4.1 Queue State Validation

Fig. 3 and Fig. 4 show the validation of queue state in the buffer for Poisson and aggregated ON-OFF arrivals. System load = 0.7 for Poisson arrival, and system load = 0.3740, 0.4082 for aggregated ON-OFF arrival. The simulation results are validated by analytical results within the 95% confidence interval. Note that we can use the same analytical method to validate simulation results with arbitrary system loads less than 1 for both arrival models. For brevity, we won't validate simulation results for other performance

analysis.

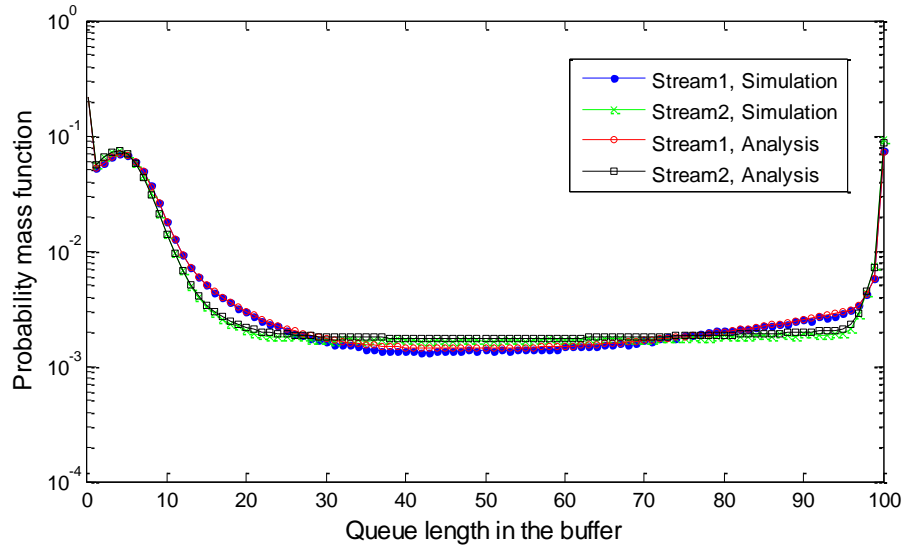


Fig. 3 Validation of queue state for Poisson arrival

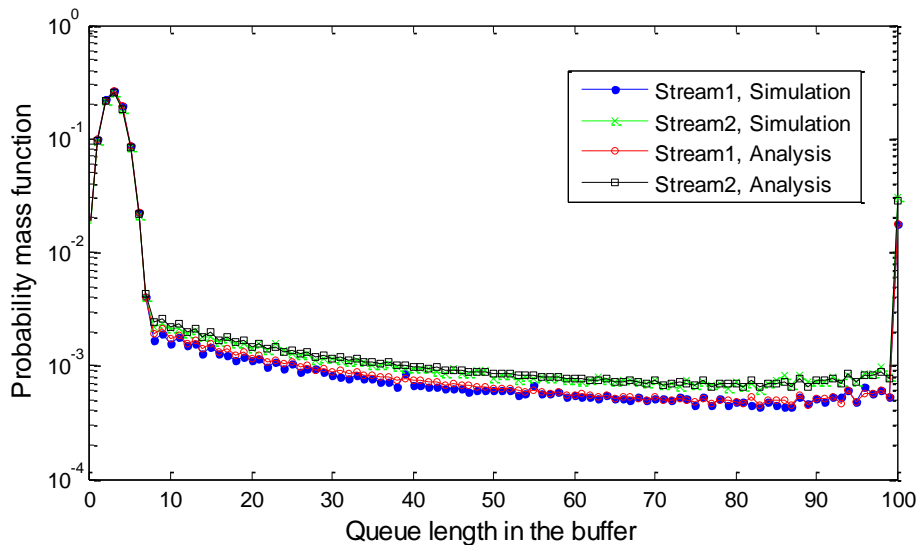


Fig.4 Validation of queue state for Aggregated ON-OFF arrival

4.2 Performance Comparisons

Fig. 5 shows the performance comparisons for (a) PDP (b) queueing delay (c) PLP (d) throughput. We have the following observations:

1. The PDPs for streams with aggregated ON-OFF arrivals are larger than the ones with Poisson arrivals with the same system loads. The difference reaches more than 4 orders in magnitude when the system load is around 0.1.

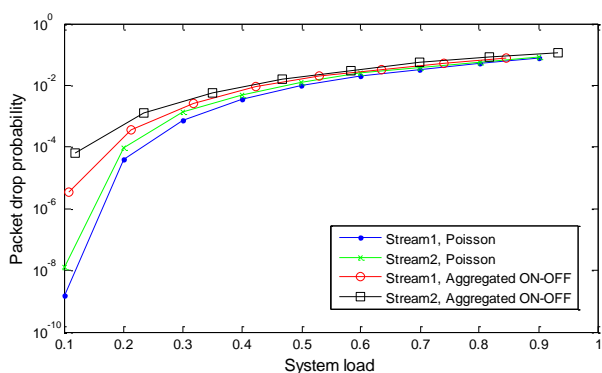
2. The average queueing delays for streams with aggregated ON-OFF arrivals are larger than the ones with Poisson arrivals with the same system loads, especially when loads are low. The difference reaches more

than 1 frame duration when the system load is larger than 0.5.

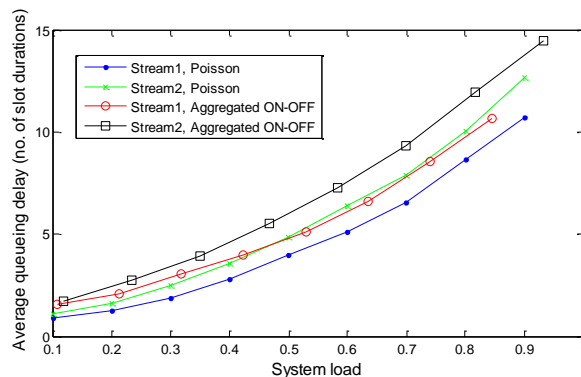
3. The PLP comparison is similar to PDP comparison except when the system load reduces to around 0.1 so that $\overline{\text{PER}}$ dominates the PLP for Poisson arrival; in other cases, PDP dominates PLP.

4. The throughput for streams with aggregated ON-OFF arrivals is smaller than the ones with Poisson arrivals with the same system loads. The difference reaches more than 0.1 packets per slot when the system load is larger than 0.8.

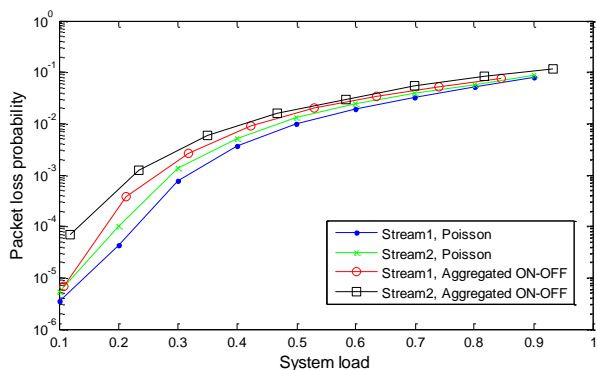
5. These comparisons give evidence of the importance of selecting proper traffic models for different traffic patterns. The aggregated ON-OFF arrival model is more appropriate in bursty service approximations.



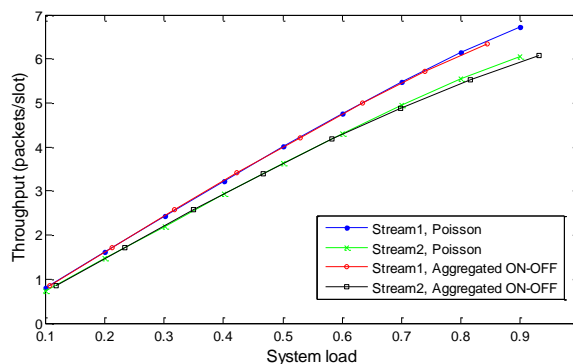
(a) PDP Comparison



(b) Queueing Delay Comparison



(a) PLP Comparison



(b) Throughput Comparison

Fig. 5 Performance comparisons for: (a) PDP (b) queueing delay (c) PLP (d) throughput

5. FUTURE WORK

In this paper, we adopt a system model which is suitable for WLAN standards, namely, the IEEE 802.11 family. More specifically, our proposed cross-layer analysis has met part of the physical layer specifications for 802.11ac such as the introduction of 256 QAM, and the system model is ready to be applied with wider channel bandwidth specified by 802.11ac with minor modifications to parameter settings for the PER analysis. In addition, a SU-MIMO scenario, which is a mandatory requirement for 802.11ac, is adopted in this paper. Some specifications for 802.11ac are not the focus of this paper, and these are left for future work.

At the data-link layer, since queueing effects are the only data-link layer factor considered in this paper, more data-link layer mechanisms could be applied into the cross-layer analysis such as ARQ. To introduce ARQ, we could use a similar method to that adopted in this paper, and a tuple with four elements, including queue state, service state, traffic arrival state and ARQ retransmission state, would need to be considered for constructing the probability transition matrix for queue state analysis. The hidden challenge would be the operation time for solving the left eigenvector of the probability transition matrix because, in such a proposed algorithm, the matrix would get much larger.

At the physical layer, 802.11ac also supports MU-MIMO in the downlink. As far as we know, in order to achieve a MU-MIMO scenario, a precoding operation such as beamforming should be applied at the AP to separate spatial streams for different users. In addition, more detectors other than the ZF detector, such as the minimum mean squared error (MMSE) detector, could be considered for equalizing at the receiver side. The challenge is in how to analytically obtain the SNR distribution for each spatial stream by ZF or MMSE detector with precoded channel gain matrix. As indicated by [35], it's quite hard, if not impossible, to obtain SNR distributions for streams with antenna correlations for MU-MIMO scenario analytically. Therefore, the use of Monte-Carlo simulation might be the only choice, although it often takes quite a long time to get reliable results.

Another interesting extension at the physical layer is to perform antenna selection for users. As specified in 802.11ac, an AP can be equipped with at most 8 antennas, while a user device can be equipped with at most 4 antennas. When there are sufficient antennas available for the AP, it can perform antenna selection for performance optimization such as to maximize the average throughput or to minimize the average PER. In addition, the available antennas for AP after antenna selection can transmit the same data to achieve diversity order or go into sleep mode to save energy. Therefore, further investigation could be focused on performance optimizations on the cross-layer analysis.

6. CONCLUSION

In this paper, a new cross-layer model combining aggregated ON-OFF arrivals and AMC with finite buffer for analyzing bursty services is proposed, and the queueing behavior for aggregated ON-OFF arrivals is analyzed, and results compared with those generated using the usual Poisson traffic model. Specifically we consider an end-to-end wireless link with multiple antennas at both transmitter and receiver side. We analyze the enlarged FSMC for the proposed cross-layer model and build up the transition matrix with a tuple of three parameters (arrival state, queue state and service state) accordingly. Our numerical results clearly indicate that the traffic characteristics have a powerful effect on system performance. This means that future studies of wireless networks with fading channels must incorporate both the fading channel and a viable model (not just Poisson) of the packet arrival process(es) in order to achieve valid performance evaluation.

REFERENCES

- [1] L.D. Servi, S.G.Finn, *M/M/1 queues with working vacations*, Elsevier Performance Evaluation, 2002
- [2] A. Doufexi, S. Armour, M. Butler, A. Nix, D. Bull, J. McGeehan, and P. Karlsson, *A comparison of the HIPERLAN/2 and IEEE 802.11a wireless LAN standards*, IEEE Commun. Mag., May 2002
- [3] Qingwen Liu, Shengli Zhou, Giannakis, G. B., *Queuing With Adaptive Modulation and Coding Over Wireless Links: Cross-Layer Analysis and Design*, IEEE Transactions on Wireless Communications, May 2005
- [4] Qingwen Liu, Xin Wang, Giannakis, G. B., *Cross-Layer Combining of Adaptive Modulation and Coding With Truncated ARQ Over Wireless Links*, IEEE Transactions on Wireless Communications, September 2004
- [5] Spyridon Vassilaras, *A cross-layer optimized adaptive modulation and coding scheme for transmission of streaming media over wireless links*, Wireless Networks, 2010
- [6] J. S. Harsini, and F. Lahouti, *Queuing with Adaptive Modulation over MIMO Wireless Links for Deadline Constrained Traffic: Cross-Layer Analysis and Design*, IEEE ICC 2007 proceedings
- [7] K. Zheng, Y. Wang, L. Lei and W. Wang, *Cross-layer queuing analysis on multi-hop relaying networks with adaptive modulation and coding*, IET Commun., 2010, Vol. 4, Iss. 3, pp. 295–302
- [8] Sheng Zhou, Kai Zhang, Zhisheng Niu and Yang Yang, *Queuing Analysis on MIMO Systems with Adaptive Modulation and Coding*, IEEE ICC 2008
- [9] J. Pitts, J. Schormans, *Introduction to IP and ATM Design and Performance: With Applications Analysis Software, Chapter 15: Resource Reservation*, Chichester: John Wiley & Sons, Edition 2, 2000
- [10] ETSI technical specification, *Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Physical (PHY) layer*, 2000
- [11] MIT 6.02 draft lecture notes, *Lecture 9: Viterbi decoding of convolutional codes*, 2010
- [12] QUALCOMM, Incorporated, *IEEE802.11ac: The next evolution of Wi-Fi Standards*, May, 2012
- [13] S. Andreev, A. Anisimov, Y. Koucheryavy and A. Turlikov, *Practical Traffic Generation Model for Wireless Networks, ERCIM workshop on Mobility*, May, 2010
- [14] L. Lipsky, M. Jobmann, M. Greiner, H. Schwefel, *Comparison of the analytic N-burst model with other approximations to telecommunications traffic*, IEEE international Symposium on network Computing and Applications, 2001
- [15] B. Chandrasekaran, *Survey of Network Traffic Models*, Informe tec, 2009
- [16] S. Shakkottai, T. Rappaport, P. Karlsson, *Cross-layer Design for Wireless Networks*, IEEE Communications Magazine, October, 2003
- [17] J. Zhang, *Bursty traffic meets fading, a cross-layer design perspective*, Proceedings of the annual Allerton conference on communication control and computing, October, 2002
- [18] E. Katsakli, A. Cateura, L. Alonso, J. Alonso-Zarate, C. Verikoukis, *Cross-layer enhancement for wlan systems with heterogeneous traffic based on DQCA*, IEEE Communications Magazine, June, 2008
- [19] H. Schwefel, *Performance analysis of intermediate systems serving aggregated ONOFF traffic with long-range dependent properties*, PhD Thesis, 2001
- [20] M. Rashid, E. Hossain, V. Bhargava, *Cross-layer analysis of downlink V-BLAST MIMO transmission exploiting multiuser diversity*, IEEE Transactions on Wireless Communications, Vol. 8, No. 9, September, 2009
- [21] T. Benson, A. Anand, A. Akella, M. Zhang, *Understanding Data Center Traffic Characteristics*, WREN'09, August, 2009
- [22] I. Elhanany, D. Sadot, *Queueing Analysis of Markov Modulated ONOFF Arrivals with Geometric Service Times*, IEEE Electrical and Electronics Engineers in Israel, 2002
- [23] X. Yang, *Designing Traffic Profiles for Bursty Internet Traffic*, IEEE GLOBECOM'02, Vol. 3, 2002
- [24] A. Zvironiene, Z. Navickas, R. Rindzevicius, *Bursty Traffic Simulation by ON - OFF Model*, Elektronika ir elektrotechnika, 2006
- [25] L. Kleinrock, *Queueing Systems. Vol. 1: Theory*. John Wiley & Sons, New York, 1975

- [26] G. Zhang, L. Sun, H. Wen, B. Wu, X. Zhu, L. Zhou, *A Cross Layer Design Combining of AMC with HARQ for DSRC Systems*, International Journal of Distributed Sensor Networks, 2013
- [27] N. Wang, T. Gulliver, *Cross layer AMC scheduling for a cooperative wireless communication system over Nakagami-m fading channels*, IEEE Transactions on Wireless Communications, Vol. 11, No. 6, June 2012
- [28] L. Musavian, T. Le-Ngoc, *Cross-layer design for cognitive radios with joint AMC and ARQ under delay QoS constraint*, IWCMC, 2012
- [29] D. Wang, L. Lu, Y. Fang, J. Wu, *Cross-layer design for LTE system with jointly AMC and ARQ on fading channel*, IEEE WCNC, 2013
- [30] B. Fu, Y. Xiao, H. Deng, H. Zeng, *A survey of cross-layer designs in wireless networks*, IEEE Communications Surveys & Tutorials, Vol. 16, No. 1, 2014
- [31] Y. Yang, H. Ma, S. Aissa, *Cross-layer combining of adaptive modulation and truncated ARQ under cognitive radio resource requirements*, IEEE Transactions on Vehicular Technology, Vol. 61, No. 9, November 2012
- [32] X. Wang, Q. Liu, G. Giannakis, *Analyzing and Optimizing Adaptive Modulation Coding Jointly With ARQ for QoS Guaranteed Traffic*, IEEE Transactions on Vehicular Technology, Vol. 56, No. 2, March 2007
- [33] J. Ramis, G. Femenias, F. Riera-Palou, L. Carrasco, *Cross-Layer Optimization of Adaptive Multi-Rate Wireless Networks Using Truncated Chase Combining HARQ*, IEEE Globecom, January 2011
- [34] G. Femenias, J. Ramis, L. Carrasco, *Using two-dimensional Markov models and the effective-capacity approach for cross-layer design in AMCARQ-based wireless networks*, IEEE Transactions on Vehicular Technology, October 2009
- [35] D. Gore, R. Heath Jr, A. Paulraj, *Transmit selection in spatial multiplexing systems*, IEEE Communication Letters, 2002
- [36] K. Jamshaid, B. Shihada, L. Xia, P. Levis, *Buffer Sizing in 802.11 Wireless Mesh Networks*, IEEE 8th International Conference on Mobile Adhoc and Sensor Systems, 2011