On the buffer occupancy of an IEEE 802.11 station in a Hot-Spot

Andrea Detti and Nicola Blefari-Melazzi

Abstract—In this paper we evaluate analytically the average occupancy of the transmission buffer of a 802.11 station (STA). The station belongs to a Wi-Fi Hot-Spot and exchanges data with a fixed host. The data exchange is regulated by the Transmission Control Protocol (TCP). The research interest is motivated by the fact that several papers analyzing TCP performance in this environment (e.g., [1],[2],[3]) assume the transmission buffer of a STA is nearly empty. Thus, the buffer occupancy of the STA is considered negligible with respect to the occupancy of the Access Point (AP) buffer located on the wireless interface. We name this assumption the empty buffer conjecture. The empty buffer conjecture strongly simplifies the modeling of the TCP send-rate. In fact, under this assumption, the delay in the WLAN is only due to packet queuing in the AP transmission buffer and packet loss occurs only in the AP. As a consequence, all TCP connections experience the same delay and packet loss probability.

However, we tested the empty buffer conjecture by means of ns2 simulations and observed a rather surprising result: in some simulation scenarios the empty buffer conjecture is not valid for some STAs.

Given these results, we felt the urge to better understand what is going on and to derive an analytical model of the system, so that fellow researchers can exploit the empty buffer conjecture only when it is right to do so. We also verified that the references quoted above ([1],[2],[3]) assume working environments in which the empty buffer conjecture is indeed valid.

Index Terms—802.11, TCP, model, buffer occupancy.

I. INTRODUCTION

We consider a Wi-Fi Hot-Spot where M stations (i.e., terminals) upload and download files to/from a fixed host by means of TCP. We analyze the average occupancy of the transmission buffer of a WLAN station (STA). The motivation of this work lies in the fact that several papers analyzing TCP performance in this environment (e.g., [1],[2],[3]) assume that the transmission buffer of a STA is nearly empty. Thus, the buffer occupancy of the STA is considered negligible with respect to the occupancy of the Access Point (AP) buffer located on the wireless interface. We name this assumption the empty buffer conjecture. The empty buffer conjecture strongly simplifies the modeling of the TCP send-rate. In fact, under this assumption, the delay in the WLAN is only due to packet queuing in the AP transmission buffer and packet loss occurs only in the AP. As a consequence, all TCP connections experience the same delay and packet loss probability.

In this section, we derive the analytical model of the average occupancy of the STA buffer. The reference scenario consists of a set of STAs having TCP connections with fixed hosts. The wireless-wired bridging is performed by an AP.

A. Assumptions and model limits

To make the model easily tractable, we make some simplifying assumptions, reported in Table I. Let us now discuss these assumptions and the ensuing model limits 1.

Assumption 1 means that the wireless part is the bottleneck. Neglecting the packet loss in the wired part seems reasonable, as loss phenomena mainly occur in the wireless part. On the contrary, the delay suffered in the wired part can not be always neglected. Our basic model exploits this assumption; however in this paper we also say how to modify the model to take into account the wired-part delay. Assumptions 2 is reasonable, since the operative system of a wireless host usually allocates a large amount of memory to its network interfaces (e.g. 1000 packets).

As regards assumption 3, the Reno version is the most widely deployed on current operative systems. So this assumption is almost always valid. Instead, when STAs use different operative systems or when users change the default configuration of the operative system the maximum congestion windows may differ. As in the case of assumption 2, we take 3 for granted to keep the basic model simple, but we also say how to modify it to account for different maximum TCP congestion windows.

Assumption 4 is the major model limit. With it, we assume that all the TCP connections start and reach a steady-state behavior. In [4] we show that this assumption is not always true. As a matter of fact, in case of heavy losses, some TCP connections may be completely starved. We are not able to capture this critical-starvation phenomenon; thus, our model is valid when the loss probability is such that critical starvation does not occur. When critical starvation does occur, the model still gives correct results, if it is possible to give as inputs to the model the number of not-starved TCP connections. Assumption 5 is quite reasonable, as the AP is the network bottleneck.

Assumption 6 saves us from evaluating the average value of the occupancy of the AP buffer, by assuming a full buffer occupancy, when losses can occur. In the no-loss case, this assumption is not necessary. The assumption is quite true as TCP flows tend to saturate the AP buffer, being this device the network bottleneck. The higher the number of connections the more verified 6 is.

As regards assumption 7, in practice, backoff procedures do not succeed in perfectly sharing the channel capacity on a per-packet basis (due to collisions and idle times). This leads to an approximation of the real performance of the system, which we will assess via simulations.

1We will make other three assumptions later on in the body of the paper, because they can be better understood only after introducing some propaedeutic concepts.
TABLE I
ASSUMPTIONS

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a1</td>
<td>The latency and the packet loss of the wired part are neglected</td>
</tr>
<tr>
<td>a2</td>
<td>The STA uplink buffer is large enough as to avoid packet loss</td>
</tr>
<tr>
<td>a3</td>
<td>The TCP version is Reno and all connections have the same maximum congestion window, W</td>
</tr>
<tr>
<td>a4</td>
<td>The packet loss probability at the AP downlink buffer is small enough: i) packet losses do not prevent the startup of TCP connections; ii) the impact of ACKs loss on the congestion window dynamic is negligible</td>
</tr>
<tr>
<td>a5</td>
<td>The AP buffer is never empty</td>
</tr>
<tr>
<td>a6</td>
<td>If the sum of the maximum congestion windows of all active connections is greater than the AP buffer size:</td>
</tr>
<tr>
<td>a7</td>
<td>The MAC layer assures a perfect per-packet fair sharing among backlogged wireless interfaces</td>
</tr>
</tbody>
</table>

B. STA buffer model

All the following analysis is based on the definition of round, which is the time interval needed to send out all the packets buffered at the AP wireless interface since the start of the round itself. The i-th round start at time $t_i$.

To derive the model, first we analyze a lossless AP buffer, then we analyze a lossy AP buffer and, finally, we combine the two treatments, obtaining a unified result. A sufficient condition for the AP buffer to be lossless [1] is that the sum of the maximum congestion windows of all active TCP connections is smaller than the AP buffer size:

$$
\sum_{i=1}^{M} (N_{dw_i} + N_{up_i}) W \leq B
$$

(1)

where $N_{dw_i}$, $N_{up_i}$ is the number of downstream (upstream) connections of the i-th STA, $W$ is the maximum TCP congestion window (in packets), $B$ is the size of AP buffer (in packets) and $M$ is the number of STAs.

1) Lossless AP buffer: let us consider a generic STA buffer during a generic round k (i.e., the round that starts at time $t_k$) and assume that at time $t_k$ the system has reached the steady state. In these conditions all TCP connections fully open their congestion windows. It follows that, when a TCP agent (sink/source) of a STA receives a packet from the AP (segment/ACK), coming from the fixed host, it generates a packet directed to the fixed host (ACK/segment) and queues it in the STA buffer. As a consequence, during the k-th round the number of packets loaded in the STA buffer is equal to the number of packets received from the AP.

In the same round, the number of packets that can leave the STA buffer is equal to the number of packets emitted by the AP, since the MAC layer assures the same transmission opportunities to all backlogged wireless interfaces. If we assume that during a round packet emissions and transmissions can be approximated with a fluid flow behavior, we can write the occupancy $Q_i(t_{k+1})$ of the STA buffer at the end of round k (i.e., at the start of round k + 1) as:

$$
Q_i(t_{k+1}) = \max \{0, Q_i(t_k) + Q_{ap}(t_k)P_{own_i}(t_k) - Q_{ap}(t_k)\}
$$

(2)

where $Q_{ap}(t_k)$ is the number of packets in the AP buffer at the start of the k-th round and $P_{own_i}(t_k)$ is the probability that at the start of k-th round a packet stored in the AP belongs to a connection of the i-th STA. Note that $Q_{ap}(t_k)P_{own_i}(t_k)$ is the number of packets transmitted by the AP to the i-th STA and $Q_{ap}(t_k)$ is also equal to the number of packets leaving the STA buffer during the k-th round, since the STA has the same transmission opportunities of the AP. The max operation accounts for the obvious fact that the buffer occupancy can not be less than zero.

In our scenario the number of active STAs is greater than one (i.e., $M > 1$). It follows that the AP has to send packets to more than one STA and that the steady-state values of $P_{own_i}(t_k)$ is strictly less than one. As a consequence, Eq.2 decreases as time increases and the average value of the STA buffer occupancy $Q_i$ is equal to zero:

$$
E[Q_i(t_k)] = Q_i = 0
$$

(3)

We can conclude that the empty buffer conjecture is indeed verified in all WLAN scenarios in which the AP buffer is lossless 2.

2) Lossy AP buffer: this case differs from the previous one in the following aspects

1) as regards downstream connections, the loss of a segment in the AP buffer reduces the TCP congestion window. Hence, the congestion window of a downstream connection is no more constant, as in the lossless case, but it depends on the packet loss probability of the AP buffer;

2) as regards upstream connections, the loss of ACKs in the AP buffer implies that, when a TCP sink located in a STA receives an ACK after a sequence of ACK losses, the TCP sink sends out a burst of TCP segments (given the fact that ACKs are cumulative). The size of this burst is equal to the number of segments cumulatively acknowledged by the received ACK. It follows that the STA may queue in its buffer more than one packet for each received ACK.

Let us first consider the downstream connections of a generic STA. Every time a segment of a downstream connection is received by a TCP sink of a STA, the TCP sink sends out the relevant ACK. This means that the STA queues a packet in the buffer at each packet reception. If we assume that, during a round, the packet emissions and transmissions can be approximated with a fluid flow behavior, we can write the number of packets $Q_{dw_i}(t_k)$ of downstream connections stored in the buffer of the i-th STA, at the end of round k (i.e., at the start of round k + 1) as:

$$
Q_{dw_i}(t_{k+1}) = \max \{0, Q_{dw_i}(t_k) + P_{own_{dw_i}}(t_k) B - \chi_{dw_i}(t_k)\}
$$

(4)

where $P_{own_{dw_i}}(t_k))$ is the probability that a packet transmitted on the wireless interface by the AP during the k-th round belongs to a downstream connections of the i-th STA

2When the TCP receiver uses the delayed-ack mechanism, queuing phenomena may occur even in the lossless case. As an example, with upstream connections only and delayed-acks, the term $Q_{ap}(t_k)P_{own_i}(t_k)$ in Eq. (2) should be corrected as $2Q_{ap}(t_k)P_{own_i}(t_k)$, since the reception of an ACK generates two TCP segments. As a consequence queuing phenomena on the STA buffer show up when $P_{own_i}(t_k) \geq 0.5$. 


and \( \chi_{dw}(t_k) \) is the number of downstream packets leaving the STA buffer during the \( k \)-th round. We also remind that in the **lossy case**, according to assumption a6, the AP buffer is always full and thus in each round the AP transmits \( B \) packets, therefore \( Pown_{dw}(t_k)B \) is the number of downstream packets transmitted by the AP to the \( i \)-th STA.

Let us now consider the upstream connections of a generic STA. We remind that, given assumption a4, the congestion window of upstream connections is equal to its maximum value, \( W \); this means that TCP’s transmission is limited by the receiver’s advertised window size, not by TCP’s congestion window size. It follows that the overall number of in-fly packets is equal to \( Nup_iW \). Given assumption a1, these packets can be in the AP buffer or in the STA buffer or they are lost at the AP buffer. Packets within the STA buffer are TCP segments, whereas packets lost or packets in the AP buffer are ACKs. This implies that, at the start of round \( k \), the number of in-fly ACKs of the \( i \)-th STA is equal to \( Nup_iW - Qup_i(t_k) \), where \( Qup_i(t_k) \) is the number of upstream packets stored in the \( i \)-th STA buffer at time \( t_k \).

Given the fact that ACKs are cumulative, at the end of round \( k \), \( Nup_iW - Qup_i(t_k) \) segments will be acked by the AP; as a consequence, the TCP senders will queue in the STA buffer the same number of segments. It follows that, at the end of round \( k \), the number of packets of upstream connections contained in the buffer of the \( i \)-th STA can be written as:

\[
Qup_i(t_{k+1}) = \max \{ 0, Qup_i(t_k) + (Nup_iW - Qup_i(t_k)) - \chi_{up}(t_k) \} = \max \{ 0, Nup_iW - \chi_{up}(t_k) \}
\]

(5)

where \( \chi_{up}(t_k) \) is the number of upstream packets leaving the STA buffer during the \( k \)-th round.

If we assume that the involved random processes are stationary, by taking the average of both members of Eqs. (4) and (5) and by approximating \( E[\max(X,Y)] \) with \( \max(E[X], E[Y]) \) we obtain an approximation of the average number \( Q_{dw}(t_k) \) of downstream (upstream) packets stored in the buffer of the \( i \)-th STA:

\[
Q_{dw}(t_k) = \max \{ 0, Q_{dw}(t_{k+1}) - \chi_{dw}(t_k) \}
\]

(6)

\[
Qup_i = \max \{ 0, Nup_iW - \chi_{up}(t_k) \}
\]

(7)

where \( Pown_{dw} \) is the steady-state probability that a packet transmitted by the AP on the wireless interface belongs to a downstream connection of the \( i \)-th STA and \( \chi_{dw}(t_k) \) is the average number of downstream (upstream) packets leaving the STA buffer during a round.

We now calculate the average occupancy of the STA buffer \( Q_i = Q_{dw}(t_k) + Qup_i \). Unfortunately, the max operator in Eqs. (6) and (7) makes this difficult evaluation. For this reason, we first evaluate \( Q_i \) when it is greater than zero and then we take into consideration what happens when the STA buffer is empty.

When \( Q_i > 0 \), Eqs. (6) and (7) may be particularized in three different ways, depending on the traffic scenarios: i) when there are only downstream connections Eq. (7) becomes \( Q_{up} = 0 \) while the max operator can be neglected in Eq. (6), since \( Q_{dw} = Q_{dw}(t_k) > 0 \); ii) when there are only upstream connections Eq. (6) becomes \( Q_{down} = 0 \) while the max operator can be neglected in Eq. (7); iii) when there are both upstream and downstream connections, the max operators of Eqs. (6) and (7) may be neglected, since both upstream and downstream connections have packets in the shared STA buffer. In the following we derive the value of \( Q_i \) in the latter case. Nevertheless, it is easy to verify that the formula of \( Q_i \) that we obtain in this case (the following Eq. (12)) will be valid also for the two cases of unidirectional data traffic.

The parameter \( \chi_{up} \) can be evaluated as the average number of packets leaving the STA buffer during a round (i.e., \( B \)) multiplied by the probability that such packets belong to upstream connections of the \( i \)-th STA. The latter probability is equal to the ratio between the average value of upstream packets in the STA buffer and the average value of all packets in the STA buffer. We now calculate the average occupancy of the STA buffer \( Q_i \) when it is greater than zero can be expressed as:

\[
Q_i = Q_{dw}(t_k) + Qup_i = \frac{Nup_iW}{(1 - Pown_{dw})} - B
\]

(9)

Now we are left with the evaluation of \( Pown_{dw} \). This probability is equal to the ratio between the number of downstream packets of the \( i \)-th STA offered to the AP buffer in the unit time and the overall number of packets offered to the AP buffer in the unit time. If we take as time unit the average packet delay in the AP buffer \( E[D_AP] \) (i.e., the average duration of a round), we have:

\[
Pown_{dw} = \frac{N_{down}NSR(p,W)}{E[D_AP] + E[D_i]} \frac{E[D_AP]}{B} = \frac{N_{down}NSR(p,W)(1-p)}{B+Q_i}
\]

(10)

where \( E[D_i] \) is the average packet delay of the buffer of the \( i \)-th STA, \( B(1/(1-p)) \) is the number of packets offered to the AP buffer in the unit time, \( NSR(p,W) \) (Normalized Send Rate) is the number of segments emitted by a TCP connection during a RTT (i.e., \( E[D_AP] + E[D_i] \)) in presence of a segment loss probability \( p \) and with a maximum congestion window...
equals to $W$. $NSR(p, W)$ is evaluated as the send-rate multiplied by the RTT; for the send-rate we use the classical expression obtained in [6] (specifically Eq. (32) of [6]) where we assume $T0=\text{RTT}$. As regards the second passage in Eq. (10), it is obtained by substituting the ratio $E[D_i]/E[D_{AP}]$ with the ratio $Q_i/B$ (this is justified by the fact that the MAC layer gives the same transmission opportunity to all backlogged transmitters).

Now, if we put in Eq. (9) the value of $Pown_{dw_i}$, as evaluated in Eq. (10), we obtain a quadratic equation in the unknown $Q_i$. By solving the equation we obtain two solutions: one is less than zero ($Q_i = -B$) and the other is:

$$Q_i = Nup_i W + Ndw_i NSR(p, W)(1 - p) - B$$

The last expression gives the average occupancy of the STA buffer when it is greater than zero. Thus, to account for the general case we must write:

$$Q_i = \max \{0, Nup_i W + Ndw_i NSR(p, W)(1 - p) - B\}$$

To complete the evaluation of $Q_i$, we need to evaluate $p$. By definition, $p$ is equal to one minus the ratio between the traffic leaving the AP buffer and the traffic offered to the AP buffer:

$$p = 1 - \frac{B}{\sum_{j=1}^{M} Nup_i W + Ndw_i NSR(p, W)\frac{E[D_j]}{E[D_{AP}]}} = 1 - \frac{B}{\sum_{j=1}^{M} Nup_i W + Ndw_i NSR(p, W) + \frac{E[D_{AP}]}{\sum_{j=1}^{M} Nup_i W + Ndw_i NSR(p, W)}}$$

(13)

If we combine the $M$ Eqs. (12), one for each of $i$ ($1 \leq i \leq M$), and Eq. (13) we get a set of $M + 1$ equations with $M + 1$ unknowns. To solve this system we resort to numerical techniques.

3) General case: To combine the lossless and lossy cases, we simply note that if we use Eq. (12) also in lossless case (i.e., $p = 0$) we get $Q_i = 0$ as given by Eq. (3). Summing up, we can consider Eq. (12) as a unique formula valid both for the lossless and for the lossy case.

Finally, we note that a rough approximation of the average buffer occupancy is given by $^8$

$$Q_i = \max \{0, Nup_i W - B\}$$

(14)

In fact, when $Nup_i W - B > 0$, downstream TCP connections experience many segment losses, which choke the congestion window and reduce the number of downstream packets making them negligible [1][2][4]. Thus, queuing phenomena in the STA buffer are essentially produced by upstream connections.

We can qualitatively explain this point by observing that the transmission of a packet belonging to a downstream connection from the AP (i.e., a TCP segment) implies the queuing of a TCP ACK in the STA buffer, while at the same time granting a transmission opportunity to the STA; this opportunity allows the STA to empty the queued ACK. On the contrary, when the AP transmits a packet belonging to an upstream connection (i.e., a TCP ACK), the STA in turns gets a transmission opportunity. However, in this case, if there are losses in the AP buffer, the received ACK may confirm more than one segment belonging to an upstream connection; thus, the TCP source on the STA may generate and enqueue in the STA buffer a number of segments greater than the transmission opportunities granted to the STA.

### III. Model Assessment Through Simulation

We test the proposed model via ns2.31 simulation. The simulation parameters are reported in Table II; parameters not reported in this table are assigned the ns2.31 default values. The simulation scenario consists of 3 STAs having TCP connections with a fixed host. The wired-wireless bridging is performed by an AP. We run simulations for different values of the number of upstream and downstream connections on each STA. In all simulations we verified that assumption a4 holds. We consider four simulation scenarios.

#### A. First simulation scenario – only downstream connections

STA n. 1 is loaded with a variable number of downstream connection; STAs n.2 and n.3 are loaded with only one downstream connection. In this case, the sum of the average congestion windows of downstream connections can not be greater than the buffer space $B$, hence Eq. (12) returns a value of $Q_i$ equal to zero. This conclusion can also be reached by looking at Eq. (9), considering that $Nup_i = 0$ and applying the max operator.

Fig. 1 reports the average buffer occupancy of the three STAs versus the number of downstream connections on STA n. 1, see curve labeled “STA n. 1,2,3 sim (only downstream).”
The model results are not plotted since they are equal to zero. The simulation results are quite near to zero, as predicted by Eq. (12). In this case the empty buffer conjecture holds.

Thus, in this scenario, all TCP connections experience the same packet loss probability and the same round trip time. As a consequence they all enjoy the same useful data rate (goodput): the WLAN system offers per-flow fairness. Therefore, the ratio between the overall goodput of STA n. 1 and that of STA n. 2 is equal to \( \frac{N_{dw1}}{N_{dw2}} \), as depicted in Fig. 2.

**B. Second simulation scenario – only upstream connections**

STA n. 1 is loaded with a variable number of upstream connections; STAs n.2 and n.3 are loaded with only one upstream connection.

Fig. 1 reports the average buffer occupancy of the three STAs versus the number of upstream connections on STA n. 1, see curve labeled “STA n. 1 sim (only upstream)” and “STA n. 2,3 sim (only upstream)”. The model results, Eq. (12), for STAs n.2 and 3 are not plotted since they are equal to zero. Instead, model results are plotted for STAs n.1. In all cases model results follow quite precisely the simulation ones. In this case the empty buffer conjecture macroscopically fails for STA n. 1. In fact, the average buffer occupancy of STA n. 1 is about equal to zero up to 2 upstream connections and then increases as \( \frac{N_{u1}}{W - B} \), when the number of connections is greater than or equal to 3. The queue of STA n. 1 can now be significantly greater than zero and this STA experiences a greater round trip time than the one experienced by the other two STAs. Therefore, the ratio between the overall goodput of STA n. 1 and that of STA n. 2 (see Fig. 2) is less than \( \frac{N_{u1}}{N_{u2}} \). In this case, the WLAN system does not exhibit a per-flow fairness, as the empty buffer conjecture would imply. This is a clear evidence of the importance of this phenomenon in real life situations, in terms of TCP performance.

**C. Third simulation scenario – upstream and downstream connections**

STA n. 1 is loaded with a variable number of downstream connections and with three upstream connections; STAs n.2 and n.3 are loaded with only one downstream connection.

Fig. 3 reports the average buffer occupancy of the three STAs versus the number of downstream connections on STA n. 1. The model results, Eq. (12), for STA n. 2 and 3 are not plotted since they are equal to zero. Model results follow quite precisely the simulation ones. Also in this case the empty buffer conjecture macroscopically fails for STA n. 1. The average buffer occupancy of STA n. 1 slowly increases with the number of downstream connections. We can conclude that upstream connections have a greater impact on the average buffer occupancy than downstream connections.

**D. Fourth simulation scenario – only upstream with a variable RTT for STA n. 1**

STA n. 1 is loaded with five upstream connections characterized by a variable round trip time; STAs n.2 and n.3 are...
This scenario is instrumental to validate our model also in the case of a fixed-network delay greater than zero. To this end, we have to modify the model as explained in footnote 5. Unfortunately, we are not able to evaluate analytically the number of packets of STA n. 1 contained in the fixed-network pipe, thus we measure this quantity by means of simulations. The results reported in Fig. 4 show that when the RTT increases, the average occupancy of the STA buffer decreases, since the number of packets contained in the rest of the network (i.e., fixed-network pipe and AP buffer) increases.

IV. CONCLUSIONS

We can draw the following conclusions:

1) if no packet loss occurs in the AP buffer, then the average occupancy of STA buffers is zero;
2) if there are only downstream connections and there are losses at the AP buffer, then the average occupancy of STA buffers is zero;
3) if there are only upstream connections and there are losses at the AP buffer, the average occupancy of the i-th STA buffer grows as $N_{up, i}W - B$;
4) when there are both upstream and downstream connections and there are losses at the AP buffer, then the average occupancy of the i-th STA buffer is the sum of two components: i) a first one due to upstream connections (and equal to $N_{up, i}W - B$); ii) a second (smaller) one due to downstream connections (and equal to $N_{dw, i}NSR(p, W)(1 - p)$).

In all cases, a rough approximation of the average buffer occupancy is given by $\max\{0, N_{up, i}W - B\}$.

REFERENCES