Layer 2 vs. Layer 3 Mechanisms for Improving TCP Performance in 802.11 Wireless LANs

Martin Heusse*‡, Guillaume Urvoy-Keller† and Andrzej Duda*
* Grenoble Informatics Laboratory, Grenoble, France
† Eurecom, Nice, France
‡ University of Colorado, Boulder, USA

***Abstract*—Access points in 802.11 wireless networks suffer from performance problems because of insufficient resources at layer 2—the DCF (Distributed Control Function) access method provides equal access probability to all devices in a wireless cell including the access point itself. Consequently, performance degrades and contention between uploads and downloads can lead to the familiar *TCP unfairness problem* [1].**

In this paper, we study the measured performance of mechanisms at different layers for improving TCP performance in 802.11 wireless LANs. At layer 2, we consider the AAP (Asymmetric Access Point) solution that keeps low buffer occupancy at the access point. At layer 3, we consider LAS-ACK, an adaptation of the Least Attained Service (LAS) policy for wireless LANs that aims at minimizing the average queue size by giving priority to the shortest connections. Using an experimental testbed, we demonstrate that AAP is a good solution especially for multimedia (delay and jitter sensitive) transfers as long as upload traffic is low. On the other hand, LAS-ACK is very efficient at minimizing the durations of most upload and download transfers as long as the distribution of flow sizes is skewed enough. The price to pay for combining LAS-ACK and AAP is the requirement of deploying LAS-ACK on all wireless stations and not only at the access point. The resulting solution actually combines the positive effects of both solutions as LAS-ACK is less sensible to the distribution of flow sizes, while multimedia flows benefit from the short queue size at the access point.

I. INTRODUCTION

We consider 802.11 wireless LANs in the infrastructure mode that in many cases become performance bottlenecks due to lower capacity of the radio channel compared to the wired part of the network. Moreover, bidirectional traffic over the half-duplex radio channel leads to unfairness problems between TCP upload and download connections and low overall performance [1].

Several authors have proposed solutions to address the TCP unfairness problem at different layers: transport, network, or MAC (cf. Section II). As the main cause of the problem is insufficient capacity of the access point, it seems promising to solve it by an appropriate solution at the MAC layer. Recent AAP (Asymmetric Access Point) proposal [2] is a pure MAC layer approach based on the Idle Sense access method [3], [4]. It privileges the access point by allocating more capacity in a dynamic way: the access point always benefits from twice the access probability of all contending wireless stations in a cell. Thus, the downlink queue at the access point never builds up unless traffic becomes intensive and downlink unbalanced.

In this paper, we compare the AAP layer 2 solution to the TCP unfairness problem with a pure layer 3 solution in which the access point gives more priority to packets from flows that have generated less traffic so far. We call the resulting policy LAS-ACK—it derives from the well-known Least Attained Service (LAS) discipline [5]. LAS-ACK is an adaptation of LAS for WLANs in which the shared radio channel is half-duplex. LAS-ACK takes into account this property by considering ACK segments in upload direction to indirectly control the rate of upload connections and enforce fairness between uploads and downlads.

We report and analyze measurement results obtained on an experimental platform implementing the proposed mechanisms. Another key aspect of our study is to consider realistic TCP workloads that mix short and long data transfers (previous studies of the TCP unfairness problem only considered long-lived connections). The main findings of the paper are the following:

- Our measurements confirm poor performance of DCF/FIFO mechanisms and show strong unfairness between uploads and downloads. This effect has already been observed experimentally, but only for long-lived connections. Our measurements show that the performance problem persists even for a mix of short and long transfers.
- Applying LAS-ACK at the access point drastically improves the response time and lowers its variability for most connections. It achieves good performance even on top of the legacy DCF MAC access method. However, it can lead to the starvation of the largest transfers, depending on the exact shape of the transfer size distribution.
- Replacing DCF with AAP helps keeping the access point buffer almost free, which is highly desirable for delay and jitter sensitive applications. However, the performance of upload traffic under AAP/FIFO degrades if its intensity becomes too high.
- Combining AAP and LAS-ACK results in reducing the sensitivity of LAS-ACK to the exact shape of the transfer size distribution, while upload traffic performance improves compared to AAP/FIFO. However, this solution requires the deployment of LAS-ACK on wireless stations, because AAP drains the buffer of the access point and nothing more can be done at layer 2 or 3.

The paper is organized as follows. Section II overviews the related work. Section IV presents the experimental platform and types of workload. Then, we report and analyze performance results in Sections V–VII. Section VIII concludes the paper.

II. RELATED WORK

Several authors have studied the TCP unfairness problem in 802.11 networks. Pilosof *et al.* proposed to modify the receiver window in TCP ACKs to pace sources on wireless stations and provide in this way more bandwidth for the download traffic [1]. Some authors proposed to solve the unfairness problem by providing a suitable scheduling mechanism at the IP layer. Eckhardt *et al.* defined an *Effort Limited Fair* scheduling for wireless networks [6]. Other authors proposed to use QoS support or service differentiation to cope with performance problems over WLANs [7], [8]. Ha *et al.* addressed the unfairness problem with two distinct queues for data segments and ACKs at the access point [9]. They tuned their relative priorities according to the number of flows in both directions and the corresponding offered window field in ACK segments.

Many authors proposed to solve the unfairness problem by using an adequate MAC access method. Leith *et al.* chose suitable parameters of IEEE 802.11e to provide fairness between competing TCP uploads and downloads [10], [11]. Other authors proposed algorithms to enhance performance in asymmetric traffic load conditions by giving more priority to the access point. However, they relied on exchanging information between the access point and wireless stations [12]–[14]. Setting up differentiation parameters is not a simple task, because the priority given to the access point needs to adapt to the current load of a cell and the number of active stations. AAP (Asymmetric Access Point) [2] sets its contention window to a constant value while wireless stations use the *Idle Sense* access method [3], [4]. In this way, AAP obtains twice transmission capacity of the sum of all active stations independently of the number of contending stations.

The related work mainly focus on improving the performance of 802.11 access points by trying to solve the TCP unfairness problem for long-lived connections. However, none of the cited papers considers the performance of a 802.11 WLAN under a realistic workload consisting of a mix of short and long flows.

III. LAS-ACK SCHEDULING DISCIPLINE

LAS (Least Attained Service), also called Foreground-Background (FB) or Shortest Elapsed Time (SET) first, is a preemptive policy that schedules the job that has so far received the least service [15]. If multiple jobs have received the same amount of service, they share capacity according to the processor-sharing policy. LAS minimizes the average response time among all work conserving disciplines not aware of the job size in advance when the job size distribution has a decreasing hazard rate [16], which is the case of many distributions including the Pareto one commonly used to model the Internet traffic. LAS favors short jobs and their impact on the response time of large jobs highly depends on the job size distribution. For the Internet traffic, it appears that in general the largest flows carry the most of the data volume. In this case, giving priority to short jobs is not detrimental to large jobs, because anyway short jobs cannot monopolize the resources.

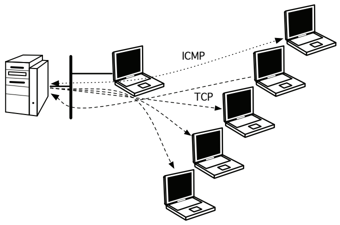


Fig. 1. Measurement platform

Any implementation of LAS requires keeping track of active flows identified by the source and destination addresses and ports. In LAS scheduling, a router identifies the first and subsequent packets in a flow, adds up the amount of data transferred by the flow, and uses this sum to insert the packet into a priority queue. Packets are served in the order of the smallest volume of transferred data first, as this corresponds to the attained service. Packets with the same volume of transferred data are served in the FIFO order.

When we want to apply LAS to schedule TCP connections in a wireless cell, we need to take into account not only data segments, but also TCP ACKs, because they make TCP sources advance in their transfers. As the 802.11 wireless link is half-duplex, both data and ACK segments contend for channel access. In the variant of LAS called *LAS-ACK*, we thus assign to an ACK segment a priority that depends on the total volume of data transferred in both directions by looking at the amount of data acknowledged by each segment. This requires maintaining two counters per TCP connection: the last ACK sequence number observed and the total data size for the bi-directional connection.

In this paper, we focus on the performance of LAS-ACK for TCP traffic only. This is well justified, because TCP accounts for most of bytes carried in the Internet in general and in WLANs in particular [17].

IV. MEASUREMENT SETUP

We use a group of six wireless stations with one acting as an access point connecting the others to the wired part of the network (cf. Figure 1). Four stations generate TCP traffic and one monitors the delay and loss rate at the access point with `ping` (we want to evaluate the performance indices that a UDP flow or TCP SYN segment may experience in given traffic conditions).

All stations may use the standard 802.11 DCF access method, switch to a different microcode running Idle Sense, or configure a fixed contention window like in AAP. This means that we can set the MAC layer of the platform to the standard 802.11 DCF or to AAP. At the packet level, the access point and stations can either use standard FIFO scheduling or LAS-ACK. In this way, we can measure four combinations of mechanisms at layer 2 and 3: DCF/FIFO, AAP/FIFO, DCF/LAS-ACK, and AAP/LAS-ACK. We have implemented LAS-ACK by modifying the BSD dummynet kernel module. We can also emulate different propagation delays over the wired part of the network in dummynet.

We use Intel IPW2915 wireless cards operating according to the 802.11a standard at the 12Mb/s rate. We have chosen this relatively low bit rate to operate in good channel conditions, but our results are still valid for higher bit rates.

We have tuned buffer sizes at layer 2 and 3. On our FreeBSD wireless stations, the default layer 2 buffer size is 64 frames in addition to the layer 3 queue of 50 packets, which makes a total of 114 packets. We have changed these default settings to 1 frame at layer 2 and 20 packets at layer 3, which corresponds to usual settings of commercial access points. A longer layer 3 buffer would increase the time spent by packets in the access point queue for DCF/FIFO or in wireless stations for AAP/FIFO. A longer layer 2 buffer has the same impact and also makes any layer 3 queueing strategies other than FIFO ineffective. More generally, large buffers are detrimental to TCP performance as they inflate the round trip time (RTT) on which depends TCP reaction to network load variations.

A. *Workload*

To observe the impact of any IP and MAC scheduling policy, the overall load on the wireless medium must be large enough. Note that relatively low bandwidth offered by 802.11 networks makes it possible to observe frequent transient overload periods even though the average utilization of the network remains small. In our experiments, we assume TCP connections arriving according to a Poisson process with rate λ adjusted such that the offered load on the wireless medium is equal to 10 Mbit/s on average, which slightly overloads the wireless link operating at 12 Mbit/s nominal rate. Since TCP controls transfers and losses can occur at the access point, the observed load is smaller than the offered load.

The workload consists of bulk TCP transfers of varying size. All TCP connections use 1500 bytes MSS. We draw the volume of data to transfer from a distribution with a fixed average value. We set the average connection size to 60 Kbytes (40 packets of 1500 bytes), which is in line with flow sizes observed on typical campus WLANs [17].

We consider two different distributions of the TCP connection size. The first one is Pareto denoted by $P(k, \alpha)$, where k is the minimum connection size and α is the exponent of the power law. The density of this distribution is given by:

$$f(x) = \alpha k^{\alpha} x^{-\alpha-1}, \qquad k \le x,\ 0 \le \alpha \le 2. \tag{1}$$

It corresponds to a realistic workload with a long tail distribution usually adopted for modeling flows in the Internet. This kind of traffic aggregates a large number of sources. We can tune its coefficient of variation (CoV—the ratio of the standard deviation to the mean) through parameter α. We have chosen the value of CoV close to 6, which is in the range of common values for WLAN traffic (e.g. observed values of CoV are between 2 and 6 [17]).

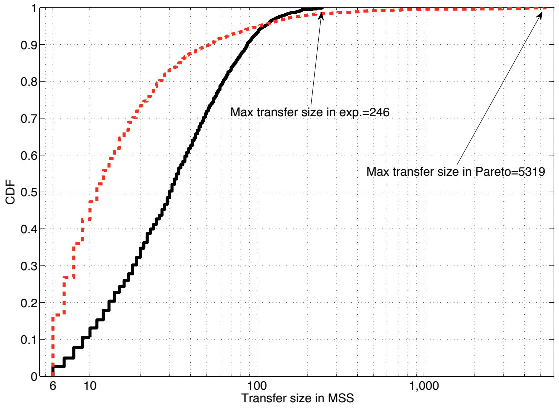


Fig. 2. Cumulative distributions functions of the transfer sizes used in the experiments.

We also consider another distribution—an exponential one of parameter μ with density function:

$$f(x) = \mu e^{-\mu x}, x \ge 0. \tag{2}$$

There are two reasons for adopting the exponential distribution, which is apparently less realistic for modeling Internet traffic than the Pareto one. First, it can correspond to traffic that does not results from aggregation of a large number of sources, which is the case for edge networks such as WLANs. Second, it represents unfavorable conditions for LAS-ACK, because LAS tends to perform better for distributions with large CoV [5] (the CoV of an exponential distribution is 1).

To account for the fact that TCP connection sizes typically have a minimum and maximum size, we set the minimum size to $k = 6$ MSS and maximum size to $P = 13,000$ MSS for both the Pareto and the exponential distributions. $13,000$ MSS correspond to a maximum transfer size of about 20 Mbytes, which is a reasonable value for a 802.11 WLAN.

Figure 2 presents the cumulative distribution function of the transfer sizes observed in our experiments. We can see that the maximum transfer size is 246 MSS for the exponential distribution and 5139 MSS for Pareto. These values are smaller than $13,000$ MSS, because we gather a finite sample for each distribution due to a limited duration of measurements.

We consider two mixtures of upload and download connections: a *symmetric* load with the same proportion of uploads and downloads and an *asymmetric* one with 66% of downloads and 33% of uploads. We consider the asymmetric load as a more realistic one. Symmetric load allows us to investigate the ability of LAS-ACK to indirectly control uploads and also corresponds to unfavorable conditions for AAP. For each combination of layer2/layer3 policy, we also consider two cases for the latency on the wired part of the path, either 20 ms or 150 ms. The first (resp. second) one is a reasonable value for a national (resp. international) path.

V. BASELINE SCENARIO: PARETO WORKLOAD

A. Preliminaries

In this section and the following ones, we present the measured performance of TCP flows over 802.11 WLANs under different layer 2 access methods and layer 3 scheduling disciplines: DCF/FIFO, DCF/LAS-ACK, AAP/FIFO, and AAP/LAS-ACK. Note that when the access point operates under the AAP method (resp. DCF), wireless stations use Idle Sense (resp. DCF). The default layer 3 policy at wireless stations is FIFO except for AAP/LAS-ACK: in this case the access point uses FIFO while wireless stations operate under LAS-ACK. This last solution requires the deployment of LAS-ACK at wireless stations, because AAP drains the buffer of the access point and nothing more can be done at layer 2 or 3.

Each experiment lasts for 100 s. Some connections are unfinished at the end of an experiment, either because of the elapsed time or an aborted transfer due to a high loss rate. We report performance results only for the connections that have completed a transfer. In the rest of this section, we focus on the results for the Pareto workload.

B. Aggregate Throughput

Figure 3 shows the aggregate throughput for all mechanisms. We can observe an increased retransmission rate of download segments for DCF/FIFO. If we look at the aggregate goodput of uploads and downloads, all mechanisms have similar efficiency at the network layer. But when we focus on the relative share of uploads and downloads for the symmetric and asymmetric cases, the advantages or drawbacks of each mechanism become visible again. DCF/FIFO favors uploads at the expense of downloads especially for symmetric traffic. DCF/LAS-ACK aims at fairly sharing resources between uploads and downloads through the way it takes into account TCP ACKs belonging to uploads. As a consequence, uploads and downloads achieve the same overall throughput under the symmetric workload. AAP based mechanisms tend to consistently favor downloads at the expense of uploads, which prevents them from achieving the same global throughput even in the symmetric case.

We also remark that the throughput graphs for DCF/FIFO do not show unfairness between uploads and downloads as pronounced as for long-lived connections. We would have expected to observe a substantially higher throughput for uploads than for downloads. But uploads do not restrain downloads, because download traffic consists of a large number of small transfers that are more aggressive than longer ones and thus the unfairness problem is less apparent in terms of throughput shares. We uncover the problem with DCF/FIFO by considering the response time of transfers below.

C. TCP unfairness under DCF/FIFO

In the rest of this section, we consider transfer related metrics, namely the *conditional connection response time*, i.e. the time required for a TCP connection of a given size to finish a transfer. We also measure *response time variability* as the difference between the 90-th and 10-th quantile of the conditional connection response time.

Figures 4 and 5 present the conditional connection response time in function of the percentiles of connection sizes for the delay over the wired part of 20 ms and 150 ms, respectively (100 percentile corresponds to 5139 MSS).

For the delay of 20 ms, we can observe that under DCF/FIFO connections last of the order of 0.5 to 2 seconds even for small transfer sizes irrespectively of the direction. When the delay is longer (150 ms), the response time of DCF/FIFO becomes even worse with significant unfairness between download and upload connections. This extends the TCP unfairness problem of DCF/FIFO to the case of the TCP workload with flows of various sizes. We can also observe high variability of the response time—Figures 4.2 and 5.2 illustrate that many connections may experience response times of up to several seconds for downloading tens of TCP segments.

D. Applying LAS-ACK

We can see in Figure 4.1 that the response time under LAS-ACK reduces to a fraction of a second except for the largest flows. Figure 5.1 shows a similar effect for the delay of 150 ms. Even more importantly for the user, LAS-ACK lowers the variability of the response time (cf. Figures 4.2 and 5.2). LAS-ACK is also able to indirectly control uploads even if the intensity of the upload traffic increases, which is the case for the symmetric load scenario (we do not present figures due to space constraints).

E. Applying AAP

Replacing DCF with AAP while still operating under FIFO shifts the point of congestion to upload stations. When the latency is significant, we observe a better response time for downloads compared to DCF/FIFO (cf. Figure 5.1–Down). Performance becomes worse for uploads, even though we note that when the latency increases, DCF/FIFO presents comparable performance of uploads to AAP/FIFO (cf. Figure 5.1–Up). Note also that with AAP/FIFO, the gain in variability for downloads is significant.

F. Combining AAP and LAS-ACK

AAP/LAS-ACK solves the performance problem of DCF/FIFO so that we can observe low variability of the response time (cf. Figures 4.2 and 5.2). All stations can now benefit from a short queue at the access point. In this combination of mechanisms, LAS-ACK improves scheduling on greedy stations, while AAP results in better sharing of wireless capacity—the wireless cell does not suffer too much from congestion.

VI. DELAY AND JITTER SENSITIVE APPLICATION

In this section, we focus on the impact of the considered layer 2 and 3 mechanisms on delay and jitter sensitive applications. We consider two types of multimedia applications that represent the current usage in the Internet: streaming transfers, e.g. YouTube video watching, corresponding to large

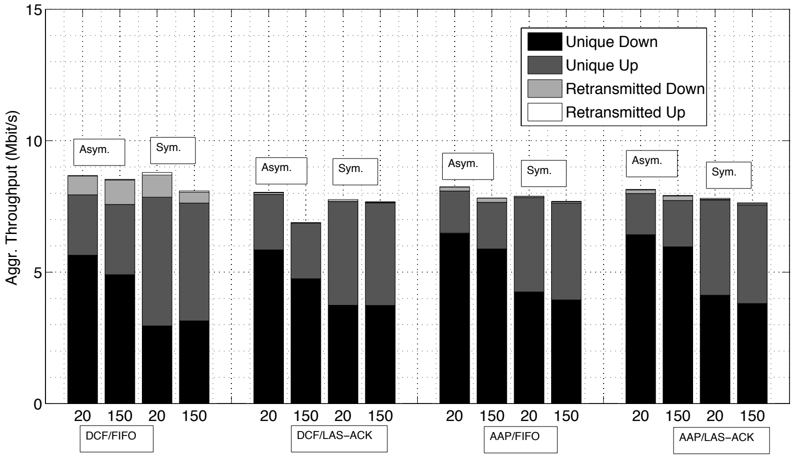
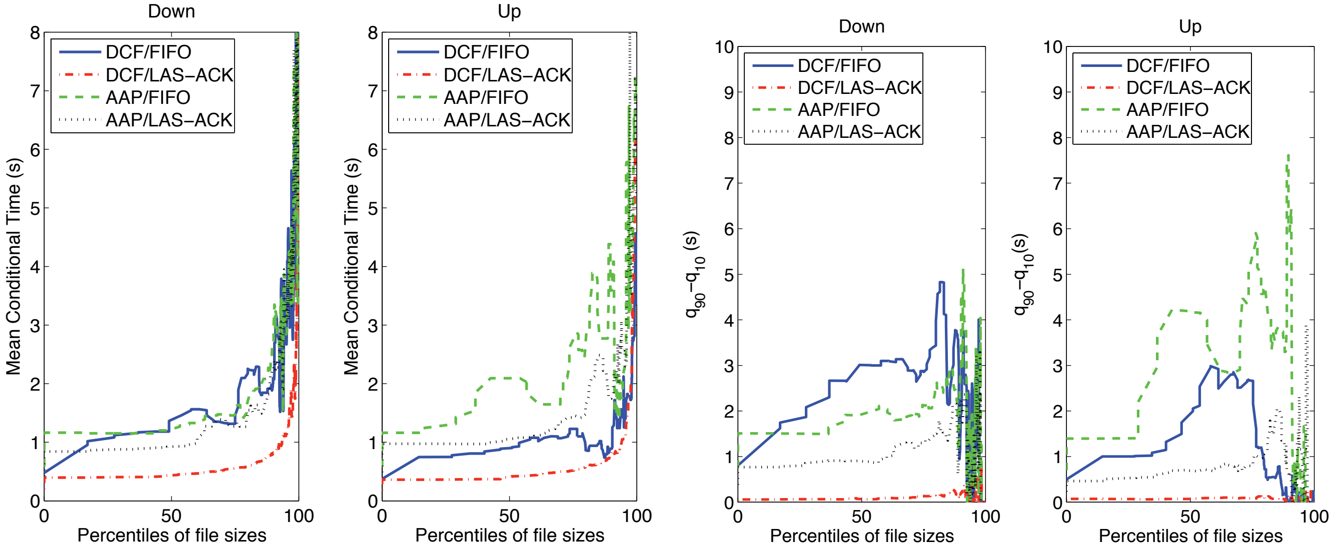


Fig. 3. Measured aggregate throughput, Pareto distributed connection size.



4.1: Conditional Response time

4.2: Difference between quantiles

Fig. 4. Asymmetric load and Pareto distributed connection sizes, 20ms delay.

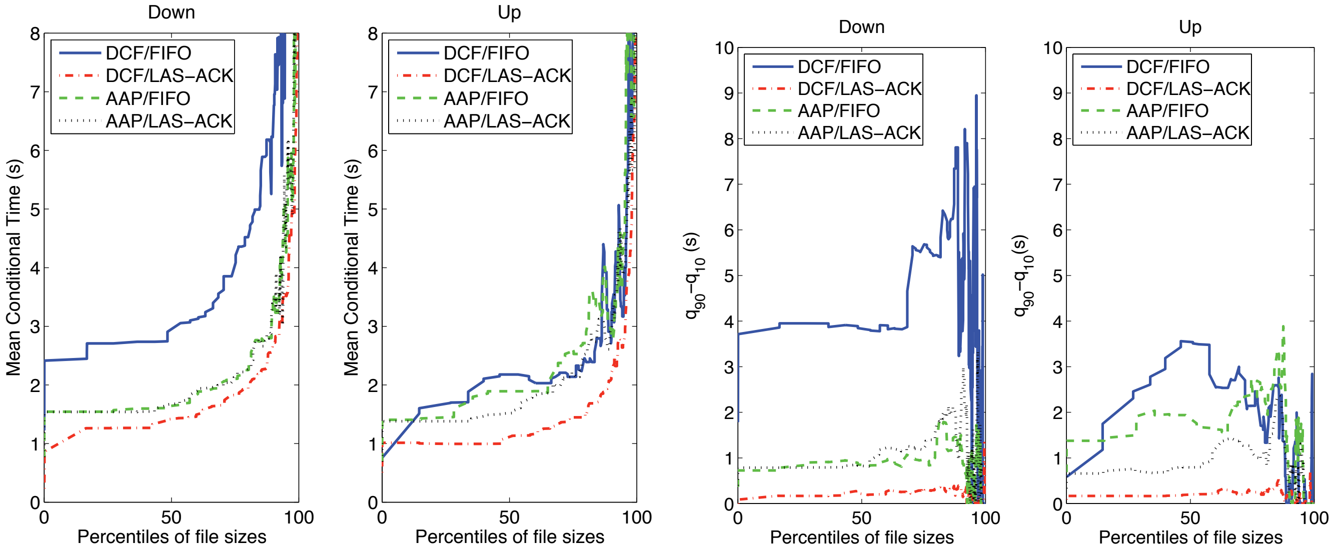
downloads and voice calls corresponding to bi-directional low rate connections.

A. *Streaming workload*

Let us first consider streaming transfers. The performance of this type of transfers directly depends on the buffer occupancy at the access point. Thus, AAP should behave the best, because it drains the access point buffer. LAS-ACK also aims at minimizing the buffer occupancy (although it operates at layer 3, so it may perform worse than AAP): it gives priority to short connections that generally do not react to losses—they do not adapt to congestion, but only delay their transmissions. Hence, by quickly serving such connections, LAS-ACK makes the buffer of the access point shorter than in the case of DCF/FIFO.

We can indirectly observe this behavior by analyzing packet round trip time (RTT) that depends on buffer occupancy at the access point over time. Figure 6 presents the distribution of packet RTT: RTT is longer for DCF than for AAP or LAS-ACK, because the buffer of the access point is full most of the time. Note however that in the case of DCF/LAS-ACK, the download rate of a large video decreases over time, because the priority of the transfer decays as its accumulated volume increases.

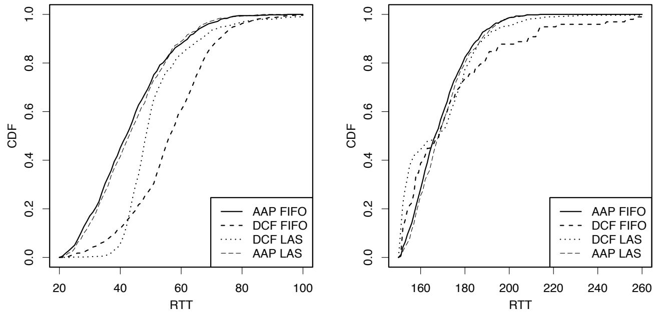
In addition to longer RTTs, DCF/FIFO also presents a drawback of prohibiting connection establishments by letting the buffer becoming full. Table I presents the loss rate for different mechanisms. We can observe a significant loss rate for DCF/FIFO and low values for all other mechanisms. Thus, DCF/FIFO can behave in an unfair manner with new connections and affect interactivity by dropping DNS requests. With its strategy to minimize buffer occupancy, LAS-ACK



5.1: Conditional Response time

5.2: Difference between quantiles

Fig. 5. Asymmetric load and Pareto distributed connection sizes, 150ms delay.



6.1: 20 ms

6.2: 150 ms

Fig. 6. Distribution of packet RTT for asymmetric load and Pareto distributed connection sizes.

TABLE I
PACKET LOSS RATE, PARETO DISTRIBUTED CONNECTION SIZES

	DCF FIFO	DCF LAS-ACK	AAP FIFO	AAP LAS-ACK
Asym., 20ms	6%	0%	0%	1%
Asym., 150ms	9%	0%	2%	1%
Sym., 20ms	14%	0%	0%	0%
Sym., 150ms	9%	0%	1%	0%

limits the probability of loosing packets. Similarly, AAP based policies quickly drain the buffer at the access point, which lowers loss rate. In summary, LAS-ACK and even more AAP are interesting solutions to provide correct performance to unidirectional (download) streaming applications.

B. *Bi-directional multimedia workload*

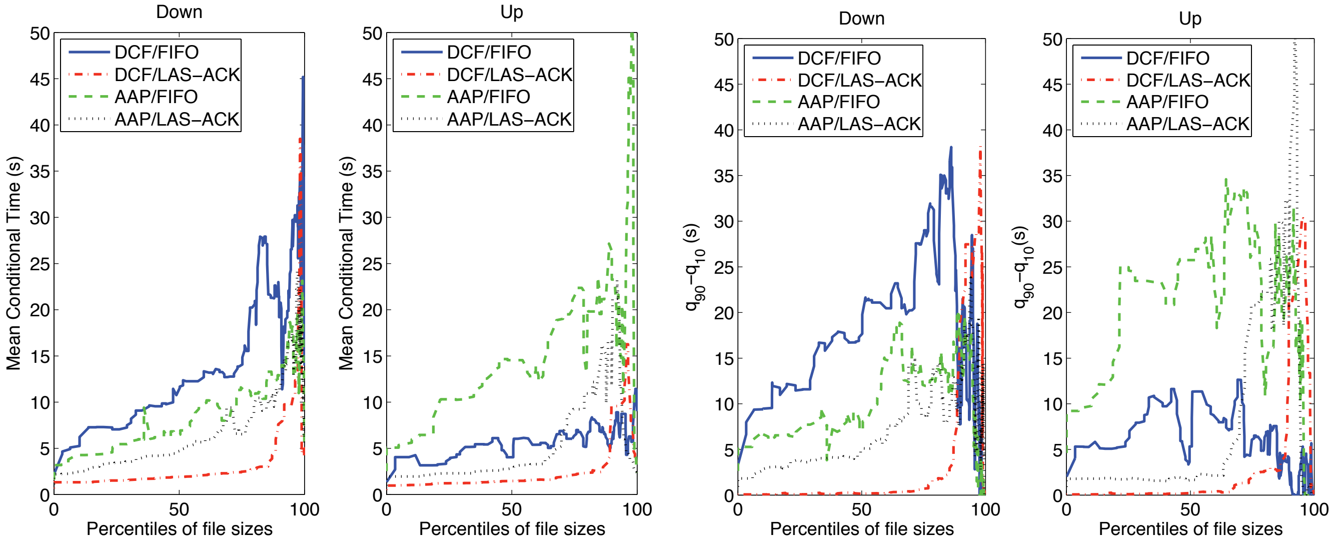
We have not specifically measured bi-directional multimedia transfers. However, we can still infer their behavior under various policies based on our previous measurements. Under DCF/FIFO, the possibly high buffer occupancy at the access point is still a major hindrance as in the streaming case. AAP/FIFO shifts congestion to the wireless station side, which solves the download problem, but may create a bottleneck for the return path, if the upload traffic intensity is too high. DCF/LAS-ACK is a good solution, if the total volume of transferred data remains low. However, even a Skype call of about 10 minutes generates around 2 Mbytes of data (3.5 bytes/s for voice calls), which corresponds to about 1300 MSS size packets. Hence, a Skype call falls in the category of large transfers from the scheduler point of view. AAP/LAS-ACK may offer a better tradeoff, but again it suffers from the limitations of LAS-ACK. Overall, it seems that AAP/FIFO remains the best solution for bi-directional multimedia transfers provided that the upload load remains low. Otherwise, one should turn to other specific mechanisms to protect multimedia flows such as for instance 802.11e.

VII. EXPONENTIAL WORKLOAD

The distribution of transfer sizes in WLANs tends to naturally exhibit high coefficients of variation due to a few very large transfers while the majority of transfers are short. Such a case is beneficial for LAS-ACK whose performance tends to improve with an increasing coefficient of variation [5]. In this section, we want to investigate the extent to which the performance of LAS-ACK can degrade when the flow size distribution has a lower coefficient of variation. We thus consider an exponential flow size distribution. While it does not seem unrealistic to observe such a flow size in the wild[1], our stance is extreme as we assume to have both an exponential flow size and high load. Due to space constraints, we will focus on the conditional response time of flows as well as on

---

[1] Imagine a set of users in a Wifi hotspot browsing the web, but not downloading large documents, and checking their mail, but not uploading nor downloading large attachments.



7.1: Conditional Response time

7.2: Difference between quantiles

Fig. 7. Asymmetric load and exponentially distributed connection size, 150ms delay.

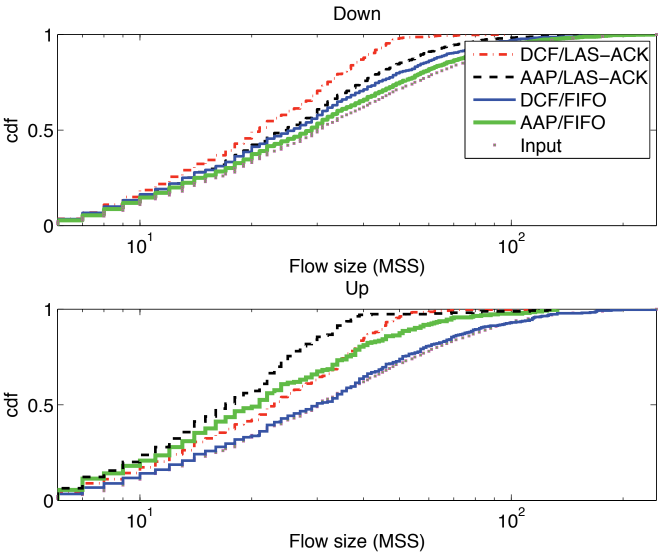


Fig. 8. Cumulative distribution functions of the sizes of completed connections for asymmetric load and exponentially distributed connection size, 20ms delay.

its variation. Results concerning the loss rate and aggregate throughput are qualitatively similar to the Pareto case.

A. *Conditional response time*

Figure 7.1 presents the conditional response times for this workload and 150 ms delay. Note that 100 percentile now corresponds to 246 MSS.

The relative performance of all four combinations is quite similar to the Pareto case, but some aspects are more pronounced. In particular, the TCP unfairness problem is more visible for DCF/FIFO. DCF/LAS-ACK still provides very low response times as well as very low variability for most of the transfer sizes. However, this comes at some cost: long connections starve under DCF/LAS-ACK, which we can see in Figure 8. It presents the cumulative distribution functions of completed connections that clearly show that LAS-ACK privileges short connections at the expense of the large ones (the input dots show the *cdf* of the transfer sizes, the same as in Figure 2). This result is in line with theoretical results on LAS showing that it is detrimental to large flows when the variability of the flow size distribution is small [15]. Note that it was not the case for the Pareto workload for which we have observed that the cumulative distribution functions of finished connections for all policies fully overlap.

We observe a similar phenomenon under AAP/LAS-ACK, although only uploads are significant here. As a conclusion, we can say that LAS-ACK continues to offer good response time to short transfers and solves the TCP unfairness problem, but at the expense of penalizing some large flows.

B. *Flow size distribution vs. traffic burstiness*

Figures 7.1 and 5.1 show that the exponentially distributed workload implies longer response times compared to the Pareto distribution. This is counter-intuitive as the queuing theory states that the average queue size of an M/G/1 queue increases with the second moment of the G distribution (cf. Pollaczek-Khinchin formula [18]). We believe that two reasons explain why the formula does not apply to the case of our experiments. First, packets belonging to the transfer of n MSS segments in our case arrive at the access point queue at a rate controlled by TCP. In contrast, if we assume that a job corresponds to a transfer in the queueing theory, the arrival of a job at the queue should correspond to all packets arriving simultaneously, which is not the case for TCP. The second reason is that TCP traffic is more bursty in the slow start phase and the exponential distribution of connection sizes tends to generate more packets corresponding to TCP connections in the slow start. Indeed, assuming no loss, the use of delayed acks, and an advertised window of 64 KB, TCP sends 128 segments before leaving the slow start phase (recall the average connection size of 40 MSS). Now, when considering the connections that are presumably in slow start (say with less than 100 MSS) in Figure 2, we observe that the ones in the Pareto workload are significantly smaller than in the exponential workload. This is normal, because the two distributions have the same mean and the Pareto workload features larger flows. As a result, we obtain that with the exponential load, 80% of the packets correspond to the slow-start phase, while it is only 40% for the Pareto case. Hence, we have more bursty traffic for the exponential case than for the Pareto workload.

VIII. CONCLUSION

The fundamental performance problem in 802.11 wireless LANs stems from the downlink packet queue that builds up when the access point does not benefit from sufficient radio channel capacity. This problem has two major consequences: it severely impacts the reactivity of short connections and interactive applications as well as results in significant unfairness between uploads and downloads.

In this paper, we have extensively evaluated the measured performance of LAS-ACK and AAP policies by considering key traffic characteristics: the skewness of the transfer size distribution, the latency of the path, the ratio of upload to download traffic, and the specific case of multimedia transfers. This study has shed light on the bad performance of DCF/FIFO that results in too large delays at high load and a very pronounced unfairness between uploads and downloads.

The two considered approaches (AAP and LAS-ACK) both attempt to drain the buffer along different angles of attack. AAP gives enough priority to the access point so that previously saturated buffer shared by stations becomes almost free. However, the bottleneck moves to stations and uplink queues build up that only a suitable queueing strategy at layer 3 can alleviate. Applying LAS-ACK at the access point decreases the queue as expected, significantly improves performance of most flows, and offers similar performance to uploads and downloads. Under extreme conditions however, we have observed that LAS-ACK can starve some of the longest flows. This effect illustrates the trade-off between unconditionally accepting flows and obtaining good performance. DCF/FIFO chooses the former option, but leads to unacceptable delays while DCF/LAS-ACK prefer to keep low response times for majority of flows.

Combining LAS-ACK with AAP results in compensating for the drawbacks of the two policies. However, the price to pay is the requirement of deploying LAS-ACK at each wireless station, as compared to DCF/LAS-ACK for which only the access point needs to support LAS-ACK.

Our study also shows the ability of AAP and LAS-ACK to protect multimedia transfers. We have distinguished between streaming (e.g. YouTube) and bi-directional (e.g. voice calls) transfers. From this perspective, both policies again outperform DCF/FIFO, with AAP offering the best performance, if the upload traffic intensity is moderate.

In the future, we intend to work on extensions of the LAS-ACK policy, alone or in combination with other layer 2 mechanisms, that could better protect multimedia transfers, while retaining the good features of LAS-ACK.

ACKNOWLEDGMENTS

This work was partially supported by the European Commission project WIP under contract 27402, the French Ministry of Research projects AIRNET under contract ANR-05-RNRT-012-01 and ELAN under contract ANR-08-VERS-008-ELAN.

REFERENCES

[1] S. Pilosof, R. Ramjee, D. Raz, Y. Shavitt, and P. Sinha, "Understanding TCP Fairness over Wireless LAN," in *Proc. of IEEE INFOCOM 2003*, vol. 2, March-April 2003, pp. 863–872.

[2] E. Lopez-Aguilera, M. Heusse, Y. Grunenberger, F. Rousseau, A. Duda, and J. Casademont, "An Asymmetric Access Point for Solving the Unfairness Problem in WLANs," *IEEE Transactions on Mobile Computing*, vol. 7, no. 10, October 2008.

[3] M. Heusse, F. Rousseau, R. Guillier, and A. Duda, "Idle Sense: An Optimal Access Method for High Throughput and Fairness in Rate Diverse Wireless LANs," in *Proc. of ACM SIGCOMM 2005*, vol. 35, no. 4, August 2005, pp. 121–132.

[4] Y. Grunenberger, M. Heusse, F. Rousseau, and A. Duda, "Experience with an Implementation of the Idle Sense Wireless Access Method," in *Proc. of CoNEXT 2007*, 2007.

[5] I. A. Rai, G. Urvoy Keller, and E. W. Biersack, "Analysis of LAS scheduling for job size distributions with high variance," in *ACM SIGMETRICS 2003, San Diego, USA*, Jun 2003.

[6] D. Eckhardt and P. Steenkiste, "Effort-limited Fair (ELF) Scheduling for Wireless Networks," in *Proc. of IEEE INFOCOM 2000*, vol. 3, March 2000, pp. 1097–1106.

[7] M. Heusse, P. Starzetz, F. Rousseau, G. Berger-Sabbatel, and A. Duda, "Bandwidth Allocation for DiffServ Based Quality of Service over 802.11b," in *Proc. of IEEE GLOBECOM 2003*, vol. 2, December 2003, pp. 992–997.

[8] D. Skyrianoglou, N. Passas, and A. Salkintzis, "Support of IP QoS over Wireless LANs," in *Proc. of IEEE VTC-Spring 2004*, vol. 5, May 2004, pp. 2993–2997.

[9] J. Ha and C.-H. Choi, "TCP Fairness for Uplink and Downlink Flows in WLANs," in *Proc. of IEEE GLOBECOM 2006*, November 2006.

[10] D. Leith and P. Clifford, "Using the 802.11e EDCF to Achieve TCP Upload Fairness Over WLAN Links," in *Proc. of IEEE WiOpt 05*, April 2005.

[11] D. Leith, P. Clifford, D. Malone, and A. Ng, "TCP Fairness in 802.11e WLANs," *IEEE Communications Letters*, vol. 9, no. 11, November 2005.

[12] X. Wang and S. A. Mujtaba, "Performance Enhancement of 802.11 Wireless LAN for Asymmetric Traffic Using an Adaptive MAC Layer Protocol," in *Proc. of IEEE VTC-Fall 2002*, vol. 2, September 2002.

[13] M. Bottigliengo, C. Casetti, C. Chiasserini, and M. Meo, "Smart Traffic Scheduling in 802.11 WLANs with Access Point," in *Proc. of IEEE VTC-Fall 2003*, vol. 4, October 2003.

[14] A. Banchs, A. Azcorra, C. Garcia, and R. Cuevas, "Applications and Challenges of the 802.11e EDCA mechanism: an experimental study," *IEEE Network*, vol. 19, no. 4, pp. 54–58, July/August 2005.

[15] M. Nuyens and A. Wierman, "The foreground-background queue: A survey," *Perform. Eval.*, vol. 65, no. 3-4, 2008.

[16] S. F. Yashkov, "Processor-sharing queues: some progress in analysis," *Queueing Syst. Theory Appl.*, vol. 2, no. 1, 1987.

[17] X. G. Meng, S. H. Y. Wong, Y. Yuan, and S. Lu, "Characterizing Flows in Large Wireless Data Networks," in *Proc. of MobiCom '04*, 2004.

[18] L. Kleinrock, *Queuing Systems, Volume I: Theory*. Wiley, 1975.