Hindawi Publishing Corporation
EURASIP Journal on Wireless Communications and Networking
Volume 2010, Article ID 818190, 16 pages
doi:10.1155/2010/818190

Research Article

Size-Based and Direction-Based TCP Fairness Issues in IEEE 802.11 WLANs

Naeem Khademi[1] **and Mohamed Othman**[2]

[1] *Department of Informatics, University of Oslo, P.O. Box 1080 Blindern, N-0314 Oslo, Norway*
[2] *Department of Communication Technology and Network, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Serdang, Selangor D. E. 43400, Malaysia*

Correspondence should be addressed to Naeem Khademi, naeemk@ifi.uio.no

Received 14 January 2010; Revised 3 June 2010; Accepted 20 July 2010

Academic Editor: Kwan L. Yeung

Cross-layer interaction of Distributed Coordination Function (DCF) of 802.11 MAC protocol and TCP transport protocol leads to two types of unfairness. In a mixed traffic scenario, short-lived TCP flows suffer from poor performance compared to the aggressive long-lived flows. Since the main source of Internet traffic is small file web transfers, this issue forms a major challenge in current WLANs which is called size-based unfairness. In addition, when sharing an access point bottleneck queue, upstream flows impede the performance of downstream flows resulting in direction-based unfairness. Proposed solutions in the literature mostly rely on size-based scheduling policies. However, each proposed method is able to solve any of these two mentioned aspects, none of them can provide both size-based and direction-based fairness in a unique solution. In this paper, we propose a novel queue management policy called Threshold-Based Least Attained Service-Selective Acknowledgment Filtering (TLAS-SAF). We show analytically and by simulation that TLAS-SAF is capable of providing both direction-based and size-based fairness and can be taken into account as a unique solution to be applied at access point buffers.

1. Introduction

IEEE 802.11 Wireless LAN has become a prevalent technology in the market due to the higher demand of users to access to the Internet from different locations. Access to the network is provided by many companies and organizations using wireless hotspots in public locations such as offices, shopping malls, airports, and restaurants as well as end users establishing local wireless networks to surf the Internet at home as a viable alternative to Ethernet connectivity.

TCP is the dominant transport protocol in the Internet, carrying typically over 80% of the bytes of a given link. The majority of applications such as peer-to-peer file sharing, web pages, email service, and podcasting have employed TCP as their transport layer protocol and it is likely to remain the transport layer of most applications in those environments in future.

Hence, fair and efficient provision of service between TCP data flows is crucial from user perspective. A number of problems due to the use of TCP in 802.11 wireless networks have been identified over the years owing to the fact that TCP has been initially optimized for wired networks.

Wireless 802.11 LANs can operate in either infrastructure mode or ad-hoc mode. In this paper, our focus is on infrastructure mode only which employs an Access Point (AP) on a wireless channel relaying traffic to and from the Internet. Wireless 802.11 MAC protocol uses two techniques which are Distributed Coordination Function (DCF) and Point Coordination Function (PCF) [1]. The main objective of DCF is to provide all WLAN stations with equal opportunity to access the transmission medium. DCF when coupled with TCP as transport-layer protocol can result in both flow level direction-based and size-based unfairness.

Cross-layer interaction of DCF and TCP leads to the unfair bandwidth allocation between downstream and upstream flows in benefit of upstream flows which is called direction-based unfairness since it deals with direction of flows. In fact, the equal opportunity of DCF makes the downstream queue at the AP a bottleneck so that data packets of downstream flows will get dropped when acknowledgment packets of upstream flows occupy the AP buffer resulting in lower bandwidth for downstream flows [2]. Another situation that causes unfairness in infrastructure-based WLANs is when long-lived and short-lived TCP flows compete for the downstream bottleneck queue at the access point. In this situation, which is a sample of size-based unfairness, packets from the long-lived flows can occupy more buffer space than the short-lived flows due to the TCP flow and congestion control mechanisms, causing higher loss for the short-lived flows and hence unfairness in benefit of long-lived flows.

Extensive research has been done to study unfairness phenomenon in WLANs. Fairness issues caused by cross-layer interaction of MAC and TCP have been studied by [3–5]. To alleviate these unfairness problems, several queue management mechanisms have been proposed which were not efficient and consistent and thus have not been facilitated by applications. This is because these solutions only solve either size-based or direction-based unfairness while even sometimes causing unfairness in another aspect. There is a need for a unique solution which can solve both types of unfairness.

The main objectives of this paper are three-fold. Firstly, we demonstrate and evaluate different types of unfairness phenomenon (short-lived versus long-lived flows as well as upstream versus downstream flows) in IEEE 802.11 WLAN. Secondly, we propose a novel queue management policy called "*Threshold-based Least Attained Service Scheduling-Selective Acknowledgment Filtering (TLAS-SAF)*" to alleviate mentioned types of unfairness to an acceptable level, and thirdly, we evaluate the efficiency and validation of our scheme using various simulation scenarios.

This paper is organized into five sections including this introductory section. The rest of the sections are as follows. Section 2 gives an overview of fairness issues in IEEE 802.11 WLANs based on different scenarios as well as different solutions proposed to provide fair and efficient bandwidth allocation among wireless nodes and cites-related works. Section 3 describes our proposed TLAS-SAF queue management policy. Section 4 describes the methodology used in this paper. The simulation framework, simulation parameters, setting and environment, assumptions, models, and performance metrics are described in this section. Section 5 discusses the performance of TLAS-SAF using simulation results. Section 6 concludes the overall research study and outlines future works.

2. Literature Reviews

This section addresses two typical unfairness issues caused by cross-layer interaction between DCF and TCP, namely, size-based and direction-based unfairness. It also reviews various solutions proposed in the literature to provide fair bandwidth allocation among TCP flows in infrastructure mode WLANs.

Fairness issues caused by TCP and MAC protocol interaction are discussed in [4–8]. Pilosof et al. [5] observed unfair allocation of the network bandwidth between upstream flows and downstream flows through network measurements. A comprehensive simulation study was also conducted to identify the causes of unfairness. The authors reported that the buffer size at the AP plays an important role in the wireless channel bandwidth allocation. They proposed a simple solution to alleviate TCP flow-level unfairness that sets the advertised receiver window of all flows to $\lfloor B/n \rfloor$ at the base station, where B is the buffer size of base station and n is the number of TCP flows. As a result, the downstream flows can achieve their fair share of the bandwidth.

Fair bandwidth allocation among short-lived flows and long-lived flows (i.e., mice and elephants) in wireless LANs has become an open issue in recent years since it has been extensively studied in wired networks [9–14]. TCP as a dominant transport protocol in computer networks and specifically in the Internet has been implemented to transfer large bulk data. Consequently, when it couples with DCF mechanism in IEEE 802.11 MAC protocol, it leads to the unfairness in bandwidth allocation among flows in benefit of long-lived flows. In fact, in wireless part, the unfairness problem between short-lived flows and long-lived flows can be more severe compared to the wired networks, since the packet loss rate and the bit error rate are usually higher owing to the nature of the wireless channel.

Several solutions to the size-based unfairness in WLANs have been proposed [3, 15, 16] which are mostly based on size-based scheduling policies operating in network and transport layer in contrast to the methods proposed in [6, 8] which rely on the MAC layer modification. The closest work to our research is [3] which employs a queue management approach to provide size-based fairness, namely, "Threshold based Least Attained Service (TLAS)". In addition, authors in [3] tried to alleviate the direction-based unfairness in WLANs based on original variant of TLAS, "Least Attained Service (LAS)" scheduling.

However LAS can solve direction-based unfairness, it results in starvation of long-lived flows and hence deteriorating the size-based unfairness. On the other hand, TLAS is able to prevent large flows to get starved and guarantee size-based fairness, but it is unable to guarantee the direction-based fairness. Subsequently, there is a need for a unique queue management solution for both types of fairness to be employed in access point buffers. Our aim in this research is to propose such a method that can provide size-based as well as direction-based fairness in one algorithm.

2.1. Direction-Based Fairness. In infrastructure-based-WLAN, access point acts as a relay between wireless and wired network. Consider a scenario based on Figure 1 in which K stations and an AP are contending for the access to the channel. Equal opportunity nature of DCF provides all stations including AP the same number of access opportunities to the wireless medium.

In this situation, each mobile station approximately gains $1/(K+1)$ share of the total transmit opportunities over a long time interval. As a result, since AP is responsible to transmit downstream data to all wireless nodes, the downstream and upstream shares of transmit opportunities would become $1/(K+1)$ and $K/(K+1)$ accordingly resulting in overall

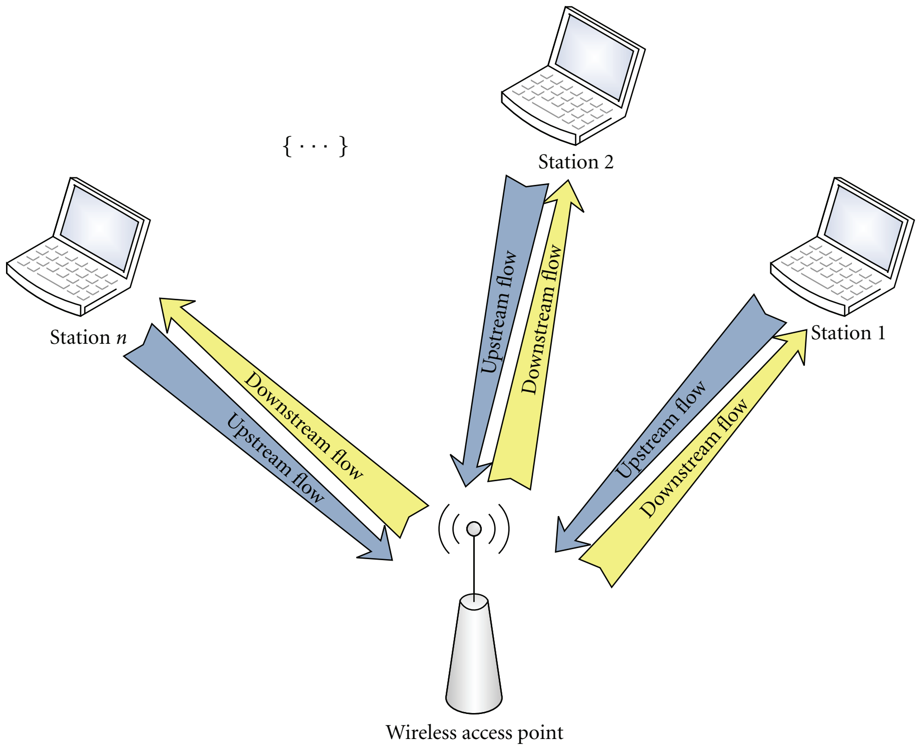


Figure 1: Scenario of AP operation.

upstream transmit opportunities to be K times more than downstream. But AP has to transmit half of the total packets in the system and hence is a bottleneck. This is owing to the fact that $1/(K+1)$ downstream transmission would also be shared by m flows. If each station transmits one downstream and one upstream flows,

$$t(\text{flow}_i) = \begin{cases} \dfrac{1}{K(K+1)}, & m_D = K;\ \text{type}(\text{flow}_i) = D, \\ \dfrac{1}{K+1}, & m_U = K;\ \text{type}(\text{flow}_i) = U, \end{cases} \quad (1)$$

where $t(\text{flow}_i)$ is the transmission opportunities for flow_i and $\text{type}(\text{flow}_i)$ can either be D or U which stand for *Download* and *Upload,* respectively. From (1), we can deduce that flows in downstream direction would suffer from lower number of transmission opportunities which leads to their lower share of bandwidth compare to the upstream flows.

In fact, DCF does not differentiate between AP and other stations in term of access to the channel making the downstream progress slowly. This happens because in presence of both types of flows, data packets of downstream flows and acknowledgment packets of upstream flows enter to the bottleneck queue due to the half-duplex nature of AP downstream queue. It means that communication is possible in both directions, but stations cannot send data simultaneously, and must be one by one. In general, the same channel is used for both transmission and reception. Consequently, the AP queue will be shared by downstream data packets and upstream ACK packets in contrast to the full duplex wired links in which upstream and downstream packets flows enter to separate queues.

When AP queue in WLAN as a bottleneck overflows, packet loss occurs. But the impact of the packet loss for downstream data packets is not the same as upstream ACK packets. Data packet loss can be detected by sender using either *triple duplicate ACK* mechanism in TCP or *retransmission timeout (RTO)*. In the first case, three duplicate ACKs will lead to the congestion window of TCP flow to be halved. In the latter case, congestion window will be set to one returning to slow-start mode. In both situations, packet loss will eventually lead to the significant throughput degradation. However, in the moment of an ACK loss, the subsequent ACK packet will acknowledge the arrival of previous one relying on cumulative nature of TCP ACKs and its loss impact on flow throughput can be seen negligible.

By this overview, we can deduce that when upstream ACK packets arrive to the AP bottleneck queue, they lead to the loss of downstream data packets and hence lower throughput for downstream flows. In contrast, existence of downstream data packet has not a considerable impact on upstream flows. In fact, AP downstream bottleneck queue plays an important role in the degree of unfairness between upstream and downstream flows (direction-based unfairness). Authors in [3, 5] found that the key parameter defining the unfairness degree is AP queue size.

Most access points in WLANs are using fixed size queues. When number of flows contending for channel bandwidth increases these queues enter to the saturation phase. In this situation, consequent drop happens for data packets as well as ACK packets which eventually lead to the lower bandwidth for downstream flows as explained before. But queue saturation never ends up owing to the fact that ACK loss does not decrease upstream flows congestion window and ACK packets will fill the buffer causing downstream data packets loss rate to get worse. To overcome this problem, AP buffer size must be large enough to accommodate all packets sending from wireless nodes. In this situation, each flow's congestion window will eventually reach to the TCP advertised receiver window which is the maximum bound of the window for each flow.

2.2. Size-Based Fairness. TCP transport protocol has been initially designed to optimize throughput for long-lived bulk data transfer. TCP congestion and flow control mechanisms have been implemented and designed in such a way that large data flows can enjoy from the best effort bandwidth allocation. While TCP *slow-start* probes the network capacity, it is followed by *steady state* and *congestion avoidance* phases. However, small files data transfers often donot reach to the steady state and congestion avoidance phases and their transfer terminate while they are in their slow-start phase. Consequently, they normally use a smaller portion of bandwidth compare to the long-lived flows since they show more conservative behavior due to the small congestion window in slow-start phase.

In infrastructure-based WLANs, when short-lived and long-lived flows share a common downstream bottleneck queue at AP, long-lived flows show a more aggressive behavior compared to the conservative short-lived flows and occupy most of the buffer space which eventually leads to the higher packet loss for short-lived flows. In fact, small file transfers are prone to the packet loss since their congestion window is too small to trigger the "triple duplicate ACK" mechanism and any loss in these flows normally leads to the "retransmission timeout" which sets the congestion window to one packet only resulting in poor performance and very low throughput. Furthermore, packet loss of short-lived flows can even become more costly knowing the fact that wireless links are prone to the bursty bit errors due to the nature of wireless channel. In fact, burstiness of errors in wireless channel transmissions may lead to the several packet losses in the same window resulting in performance degradation of TCP flows. All these issues cause that short-lived flows suffer from higher variability of transfer time compared to the long-lived flows.

Studies have revealed that Internet traffic shows a high variability property and most of the TCP flows are short, while more than half of the bytes are carried by less than 5% of the largest flows [17, 18]. Since Web data transfer is the most popular service in the Internet, small files are expected to comprise the main fraction of the Internet traffic. In contrast, the largest flows seen in the Internet are originated from peer-to-peer file sharing services that have become popular in the recent years. The high variability of Internet traffic can negatively affect the end user experience in general due to the nature of TCP transport protocol and FIFO scheduling mechanism used in most network routers. In WLANs, the situation is even worse because of the reasons mentioned previously. While most of the traffic flows triggered by user interactions are short-lived flows, size based unfairness issue in WLANs leads to the degradation of interactivity experienced by end users.

Several solutions have been proposed to solve the problem of size-based unfairness in IEEE 802.11 WLANs mostly relying on size-aware scheduling policies [3, 15, 16]. A size-aware scheduling policy refers to the queue scheduling scheme that differentiates and prioritizes packets based on their corresponding flow size. Mentioned proposed solutions normally use a class of size-aware scheduling policies which gives service priority to packets from short flows to improve overall user perceived performance without penalizing the performance of long flows too much.

2.3. Related Works. In this part, we will study various methodologies proposed to solve any of size-based or direction-based unfairness in 802.11 WLANs. These are Advertised Receiver Window Setting, Dynamic Buffer Sizing and different variants of size-aware scheduling policies (e.g., LAS and TLAS). Furthermore, we will explain that none of these works are able to solve both types of unfairness.

Advertised Receiver Window Setting. AP buffer size needed for TCP fairness is estimated by [5] as

$$B \geq (\partial M W + N W), \tag{2}$$

where ∂ is the number of ACK packets per data packets in TCP, W is the TCP-advertised receiver window for all flows, and N and M refer to the number of downstream and upstream flows, respectively. A more accurate estimation of minimum buffer size to achieve fairness in the scenario of one TCP upstream flow and several TCP downstream flows is estimated by [19] as

$$B \geq W + 4N\sqrt{\frac{0.38\left(W^2 + 2W\right)}{3}}. \tag{3}$$

From (3), the work [20] infers that the maximum value of the advertised receiver window associated with AP buffer size ensuring fairness is calculated as

$$W_B = \frac{-2N^2 - 3B + \sqrt{4N^2 + 12N^2B + 6N^2B^2}}{2N^2 - 3}. \tag{4}$$

In real infrastructure-based WLANs, AP is normally connected to the wired link and provides Internet service via that link. Thus, data is in transit in the link proportional to the link delay. Consequently, amount of total buffer including the data in transit in the link and AP buffer increases proportional to the increase of link delay. Considering data in transit in the link, the maximum value of the advertised receiver window ensuring fairness is calculated by [20] as

$$W_D \leq W \leq W_B + W_D, \tag{5}$$

where W_D is the TCP window size associated with link delay and calculated as

$$W_D = \frac{\left(\text{delay}_{\text{wired}} + \text{delay}_{\text{wireless}}\right)\text{thr}_{\text{total}}}{4(N+1)\text{packet\_size}}, \tag{6}$$

where $\text{thr}_{\text{total}}$ is the total achievable throughput on the channel. Both TCP fairness and total throughput are ensured as long as advertised receiver window size is the value within the range of (5). Proposed method in [20] guarantees per-flow fairness among TCP flows while maintaining maximum achievable throughput. From (3), we can also infer that as the number of downstream flows (N) increases, the needed buffer size B increases in proportion with N implying the high impact of number of downstream flows on needed buffer size to assure fairness.

Dynamic Buffer Sizing. In contrast to the static buffer sizes employed in mentioned methods, in the work [21] proposed a dynamic buffer sizing method to achieve flow level fairness in IEEE 802.11e WLANs. In proposed method by [21] adaptive buffer size to achieve fairness is calculated by

$$B = \min\left(\frac{T}{T_{\text{serv}}} + a, B_{\max}\right), \tag{7}$$

where T is the target queuing delay which should be guaranteed to each packet and $T_{\text{serv}}(t)$ is the interservice time of queue at time t and a is the overprovisioning amount to accommodate short-term fluctuations of AP queue service rate. Authors in [21] proposed T_{serv} to be calculated by

$$T_{\text{serv}} = \alpha T_{\text{serv}} + (1 - \alpha)(t_e - t_s), \tag{8}$$

where t_s and t_e refer to the arrival time of packet to the head of queue and successfully receive time of MAC ACK of that particular packet, respectively. Using this method, AP buffer size adapt its size in such a way that can provide fair bandwidth allocation among upstream and downstream flows.

Least Attained Service Scheduling. LAS is a preemptive scheduling policy that favors short jobs without prior knowledge of the job sizes. To this end, LAS gives service to the job in the system that has received the least amount of service so far [22]. At any given moment, the set of jobs which have received the least service share the processor (or any resource). A newly arrived job always preempts the job currently in service and keeps the processor until it gets finished, or until the next arrival occurs, or until it has achieved an amount of service equal to that received by the job preempted on arrival, whichever occurs first. In the past, LAS was believed to heavily penalize large jobs. However, it has recently been proved that the mean response time of LAS highly depends on the job size distribution [23].

The concept of *job* has been employed by processor sharing policies as a piece of workload which arrived to a system at once. However, in computer networks this definition might not be correct since a flow of packets does not arrive at once to a network resource. In fact, in computer network a *flow* is consisting of a sequence of packets arriving in timely manner that are multiplexed with other flows. Consequently, analytical modeling of LAS for jobs cannot simply apply to flows. Instead, we use simulation techniques to study the performance of LAS in network flows.

Based on LAS definition in network flows, the next packet to be served is the one that belongs to the flow that has received the least amount of service so far. Consequently, LAS will serve packets from a newly arriving flow until that flow has received an amount of service equal to the amount of least service received by a flow in the system before its arrival. If two or more flows have received an equal amount of service, they share the system resources fairly. In a network priority queue LAS, as a queue management policy, operate as follows. When a packet arrives to a full queue, LAS first inserts the arriving packet at its appropriate position in the queue (based on the service it has received so far) and then drops the packet which is at the end of the queue [3]. Therefore, LAS gives buffer space priority to short-lived flows as the packet discarding policy under LAS biases against flows that have utilized the network resources the most.

Studies have shown that LAS significantly reduces the mean transfer time and loss rate of short TCP flows as compared to DropTail first-in first-out (FIFO) scheduler, with a small increase in mean transfer time of large flows. They also show that under moderate load values a large flow under LAS is not starved when competing with short TCP flows. In contrast, the performance of long-lived flows under LAS deteriorates severely when competing at high load with short flows.

LAS scheduling policy has been proposed to be employed by network routers recently [13, 23]. A flow is identifiable at the network router by its source and destination addresses and ports. To implement LAS, router employ a counter for each flow to keep track of the service that flow has received so far. When a new packet arrives at the network, router compares its corresponding flow's counter to other counters and insert the packet in the queue position according to its flow received service in a sorted way.

A simpler implementation of LAS can employ TCP packet sequence number as the amount of service which its flow has received so far since TCP sequence numbers are associated with the number of received packets at the destination and can be taken into account as achieved service of flow. In this situation, priority queue is sorted based on packet sequence numbers. When a new packet arrives at the queue, it will be inserted at its appropriate position based on its sequence number and finally in case of full queue, the packet with maximum sequence number (at the tail of queue) will be discarded. In the situation when no sequence number more than packet sequence number is found in a full queue, the arriving packet will simply get discarded since it should normally belong to a flow which has received the maximum service compared to other flows so far. Finally, when all flows received the equal service to that flow, arriving packets from it can be served in queue.

One drawback of LAS scheduling scheme is that one newly arriving long-lived flow can block all other existing long-lived flows until the time it receives the same service they achieved. In this case, all other long-lived flows get starved which causes severe unfairness among long-lived flows. *Threshold-based least attained service* (TLAS) scheduling has been proposed to solve this type of unfairness. Its main idea is to give the newly arriving flow service priority up to a certain threshold (e.g., 50 packets). Once the threshold is reached, FIFO scheduling is employed on this flow [3]. Authors in [3] showed by simulation that TLAS can guarantee fairness for short-lived flows as well as long-lived flows.

Since LAS (and also TLAS) gives highest priority to the first packets of each flow, it prevents these packets to wait in the queue as they experience little or no queuing delay resulting in decrease of RTT for first packets of flow. Since the RTT dominates the transmission time of short flows, reduction in RTT results in reduction of transmission time of short flows [22]. In the presence of congestion in slow-start phase, packet loss will eventually trigger the retransmission timeout (RTO) because congestion window is too small to trigger the triple duplicate ACK mechanism. Since the computation of RTO depends on RTT sample in TCP protocol, RTT_{LAS} will be smaller than RTO_{FIFO} owing to the fact that RTT_{LAS} is smaller than RTO_{FIFO} as we previously mentioned. This leads to reduction of transfer time in LAS policy compared to FIFO even in the presence of congestion.

LAS as a Solution for Size-Based and Direction-Based Unfairness in 802.11 WLANs. In addition to the wired network routers, LAS scheduling variants have been proposed to be used in infrastructure-based wireless local area networks so as to solve the size-based and direction-based unfairness problem at the AP router as WLANs suffer from the poor performance of short-lived and also downstream flows [3, 16, 24].

In wired networks where links are in general full duplex, LAS is applied to each direction of the link independently from each other. In the case of 802.11 WLANs where wireless channel is half duplex, in the work [24] proposed to apply LAS on a connection basis which means that the priority of a packet at the AP is based on the total amount of traffic sent by the corresponding connection. For the case of TCP acknowledgments, packet priority will be set equal to the amount of bytes carried by the data packets on the other direction of the connection.

Two different solutions based on LAS queue management policy are proposed by authors in [3] to solve size-based and direction-based unfairness in 802.11 WLANs. For size-based unfairness, they employed TLAS with threshold of 50 packets to prevent long-lived flows to get starved while guaranteeing better performance for short-lived flows. In addition, they used LAS to solve direction-based unfairness by achieving fair bandwidth allocation among downstream and upstream flows. However, these two variants of least attained service scheduling could solve both types of unfairness independently from each other, none of them can solve both unfairness problems as a unique solution.

In fact, using TLAS to achieve size-based fairness will not guarantee direction-based fairness since TLAS gives the equal service to all flows until a certain threshold which normally is a small value and it behaves as FIFO for the rest of transmission period resulting in direction-based unfairness for most part of the transmission period (specifically for long-lived downstream flows). On the other hand, while using LAS to provide direction-based fairness, current long-lived flows get locked by arrival of a new long-lived flow. In addition, when the traffic workload is too high and AP buffer is normally saturated, applying LAS will lead to the penalizing of long-lived flows resulting in their higher variation of transfer time.

Table 1: Comparison of WLAN-oriented solutions.

Method	Size-based fairness	Direction-based fairness	Comment
LAS [3, 16]	No	Yes	Starvation of long flows
TLAS [3, 16]	Yes	No	
Advertised window setting [5, 20]	No	Yes	
[6, 8]	No	Yes	MAC modification/802.11e
[11]	Yes	No	
[24]	No	Yes	
Adaptive dynamicBuffer sizing [21]	No	Yes	

Since implementation of two different queue management policies in AP buffer of WLANs is not possible, there should be a unique solution to guarantee size-based and direction-based fairness among competing TCP flows to be applied at the physical AP infrastructures. Our aim in this research is to propose such a method in which both types of fairness in 802.11 WLANs can be assured. We call this method as *Threshold-based Least Attained Service-Selective ACK filtering* (TLAS-SAF) which barrows its idea from two different queue management policies studied in [3], namely, TLAS and *Selective Packet Marking-ACK Filtering* (SPM-AF). The closest work to our research is [3] and we develop our methodology based on the concepts practiced and evaluated by this work.

Comparison of Proposed Methods in the Literature. In order to evaluate and compare the WLAN-oriented solutions proposed in previous sections we have categorized them as in Table 1. It is obvious that none of these solutions could solve both typical types of unfairness in 802.11 WLANs.

3. TLAS-SAF Queue Management Policy

In Section 2.3, we explained with detail that TLAS can improve performance of short-lived flows without deterioration of long-lived flows (size-based fairness), but it is unable to guarantee direction-based fairness. On the other hand, LAS can provide fair bandwidth allocation among downstream and upstream flows (direction-based fairness) but it leads to the starvation of long-lived flows. Hence, there is a need of a unique algorithm to solve both types of problems in one scheme.

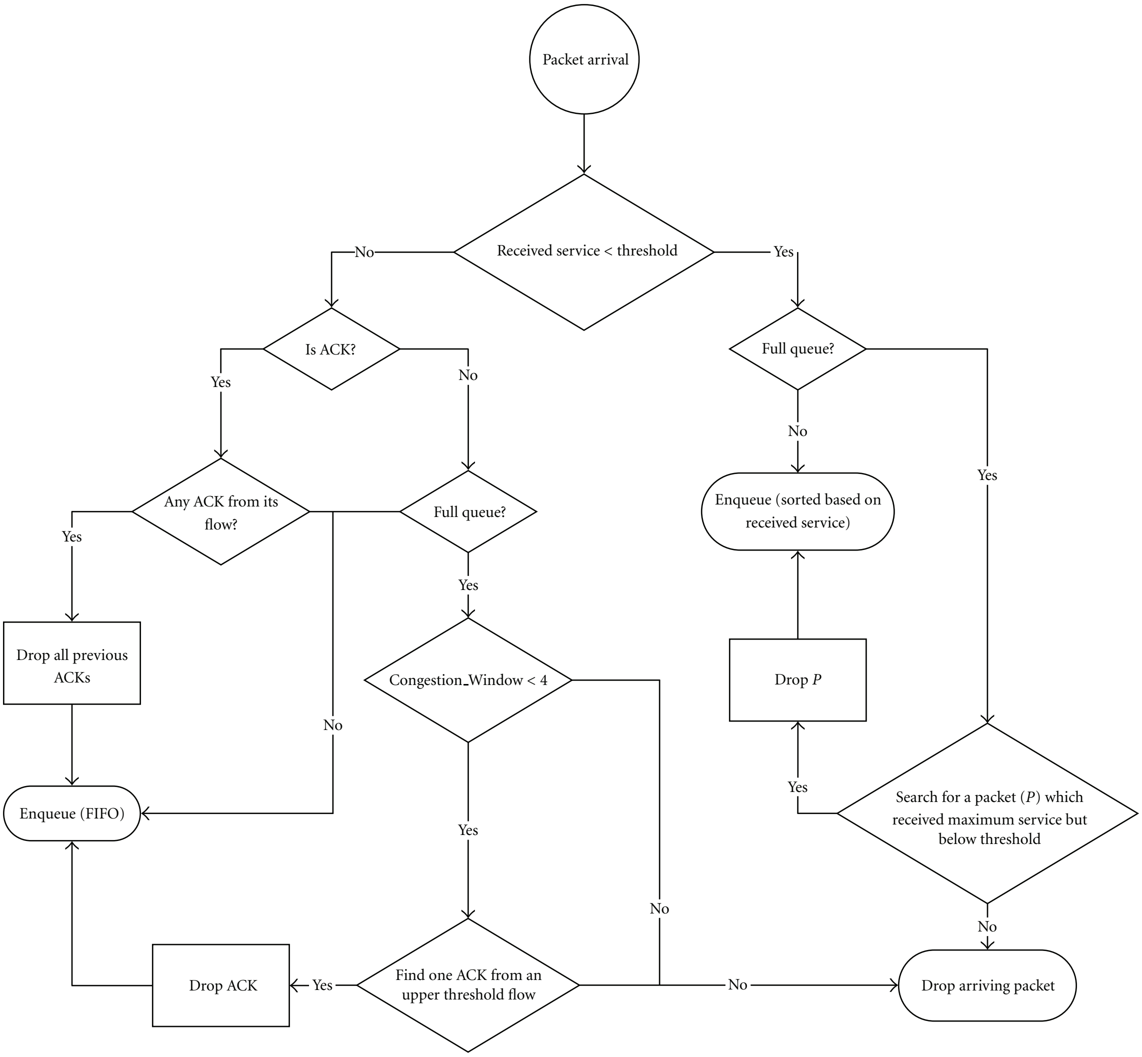


FIGURE 2: TLAS-SAF flow chart.

To achieve this solution, we propose a novel queue management policy called *Threshold-based Least Attained Service-Selective ACK Filtering* (TLAS-SAF). This algorithm barrows its main idea from two different queue management policy, namely, Threshold-based Least Attained Service (TLAS) and Selective Packet Marking-ACK filtering (SPM) [3, 25]. It sets a minimum threshold for service that should be guaranteed for every network flow and it behaves the same as LAS with all packets below that threshold. Once service threshold is reached for a flow it behaves with flow as SPM-AF (specified in [3]) does. Consequently, TLAS-SAF inherits TLAS capability to provide size-based fairness while employing SPM-AF characteristics for larger flows to provide direction-based fairness as well.

TLAS-SAF works as follows: it firstly gives service priority to newly arriving packets of a flow until a certain threshold (e.g., 50 packets). Once threshold is reached, FIFO scheduling will be imposed on this flow. If packet of a flow below threshold encounters to a full buffer, TLAS-SAF looks for a packet of a below-threshold flow which has received the maximum service so far (similar to TLAS) and discards it. But if received service for such a flow was less than received service of arriving packet flow, the arriving packet will be discarded instead.

However, some data packets in network flows are more crucial than others since their loss result in retransmission timeout (RTO) while loss of other data packets will trigger *fast retransmission* mechanism. Authors is [3] show that if any data packets loss happens when the congestion window is smaller than four packets, it leads to the coarse-grained RTO because congestion window is too small to trigger *triple duplicate ACK* mechanism to *fast retransmit* data packet. In this case, TLAS-SAF gives priority to data packets which their congestion window is smaller than four for the flows have already passed the threshold. In full buffer situation, if any new data packet with corresponding congestion window smaller than four arrives and its flow has already achieved the threshold service, it will be admitted in FIFO order at the queue but an ACK packet from above threshold flows will be discarded from the queue.

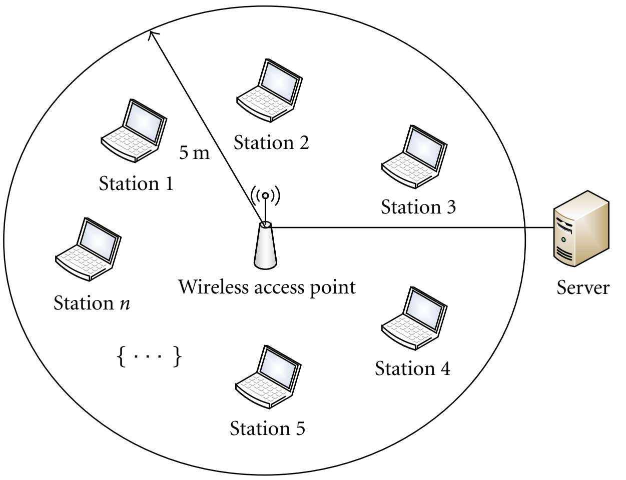


Figure 3: Network topology used in ns-2 simulations.

In legacy TCP protocol as well as most of the other TCP variants, the value of congestion window is only available at the TCP sender. To provide the availability of per-flow congestion window value to AP, a slight modification of TCP protocol is required. In our ns-2 implementation, we modified the TCP header to carry the congestion window of its corresponding flow. Once a packet is admitted at AP queue, it retrieves the congestion window value from packet header. One possible solution for real test-bed experiments is to use the unassigned *options* bits in TCP packet header to carry the congestion window value. This can be further implemented in TCP source code in Linux kernels. By the way, for this paper, our research scope is limited to the simulation only.

In fact, loss of ACK packets in a flow has negligible impact on performance of network flows relying on cumulative nature of ACK packets since the next ACK packet for that flow will supersede the information in the previous ACKs. In addition, TLAS-SAF treats ACKs in such a way that at any given moment only one ACK from a certain upper threshold flow exists in queue providing more buffer space for data packets. For these flows when an ACK arrives to a full buffer, TLAS-SAF checks whether any other ACK from its flow exists in queue and if found it will simply get discarded. TLAS-SAF *does not* discard any ACK or data packet of flows that have not reached to the threshold to accommodate upper threshold packets providing better performance and response time for short-lived flows. Figure 2 shows a schematic flowchart of TLAS-SAF.

Calculation of current received service in TLAS-SAF is quite straightforward. TLAS-SAF assumes that each flow associated with a newly arrived packet has already received an amount of service corresponding to its packet sequence number. Thus the priority value of each packet is set to its sequence number. However, for ACK packets, the priority is set to the sequence number of the last data packet it acknowledges. In ns-2, ACK packet sequence numbers are set to the data packet sequence numbers they acknowledge and ACK flows donot have different sequence series than their corresponding data packets, making calculation of received service of a flow much simpler.

Another important issue in TLAS-SAF is to choose an appropriate value for *threshold* to optimize the performance of this algorithm under Internet traffic flows. Threshold value selection in TLAS has been studied in [26, 27]. These works mention that a good threshold value is the one that is able to capture all short flows and therefore its selection depends on traffic pattern that TLAS is operating in. Small threshold values result in a behavior more similar to DropTail/FIFO while large values of threshold lead to a behavior more similar to LAS and therefore are not favored due to the starvation of large flows. Since Internet traffic flow sizes are highly variable and majority of flows are small size web data transfers we choose to set the threshold value at 50 packets to cover the majority of short flows. This value is used for both TLAS and our proposed TLAS-SAF simulations.

4. Research Methodology

In this section, we explain about all aspects of our research methodology such as simulation tool, framework, settings and environment, and parameters being used to evaluate the performance of TLAS-SAF. In addition, simulation model, network topology, assumptions, and performance metrics are explained with details in this section.

To evaluate the performance and validity of our TLAS-SAF queue management policy, we compare its performance with the traditional FIFO/DropTail buffer management scheme as well as other proposed variants of size-based scheduling policies, namely, LAS and TLAS through extensive simulations in ns-2 [28]. Network Simulator-2 (ns-2) is a popular discrete event simulator which has been taken into account as a de facto standard tool for networking research among academicians. To support these queue management policies, we extend ns-2 by implementing new C++ classes.

Figure 3 shows network topology used for ns-2 simulations in this research. There are n number of wireless nodes and one wireless station as Access Point (AP) in an infrastructure-based IEEE 802.11 WLAN. A server is connected to AP via a wired link. All wireless stations exchange data to and from a server via AP. All nodes are in equal range of $5m$ from AP.

Our set of simulations constitutes of two parts. First part belongs to the direction-based (upstream versus downstream flows) performance evaluation and validation of our proposed TLAS-SAF scheme compared to the other queue management policies (DropTail, LAS, and TLAS). The second part relates to the size-based (short-lived versus long-lived flows) performance evaluation of our proposed scheme compared to the other queue management policies. We study our proposed scheme under these two sets of simulations to prove that it can improve the level of fairness for both size-based and direction-based aspects.

Table 2 shows the common simulation parameters for both direction-based (upstream versus downstream flows) and size-based simulations (short-lived versus long-lived flows). IEEE 802.11b MAC protocol has been used in all

Table 2: Common simulation setup parameters for all simulations.

Examined protocols	TLAS-SAF, TLAS, LAS, DropTail
Simulation area	670 × 670 m
Propagation model	Two-ray ground
Traffic type	TCP
MAC layer	802.11
Antenna	Omni
Node to AP distance	5 m
Threshold (TLAS, TLAS-SAF)	50 packets
802.11 MAC setting	
MAC version	802.11b
Slot time	20 μs
SIFS	10 μs
Preamble length	72 bits
Short retry limit retransmissions	7
Long retry limit retransmissions	4
PLCP Header Length	48 bits
MAC basic rate	1 Mbps
MAC data rate	11 Mbps
CWMin	31
CWMax	1023
TCP setting	
TCP protocol	TCP NewReno
Packet size	1460 bytes
Segment per ACK	1 (ns-2 default)
Application	FTP
AP to Server link	
Bandwidth	100 Mbps
Wired delay	2 ms
Queue policy	DropTail

of our simulations. MAC parameters are specified based on IEEE 802.11b MAC specifications. TCP NewReno variant has been used as our preferred transport protocol for our evaluations since this variant is popular in current trend of computer networks. FTP application is used to generate TCP traffic for large/small files as well as upload/download data transfers. Traffic is relayed to and from server by AP via 100 Mbps wired link with wired delay of 2 ms.

4.1. Size-Based Fairness Evaluation. To evaluate size-based fairness n number of wireless nodes is defined which is varied from 6 to 46 nodes. Half of this wireless stations transfer large files with a fixed file size of 1000 data packets (packet size equal to 1460 bytes) using FTP connections. Each of these stations downloads 10 such large files sequentially. The other half of the wireless stations generate small 10 packets file size download traffic with exponential on-off distribution of interarrival times with burst time of 500 ms and idle time of 500 ms and with the same packet size as the large files. The small file transfers are repeated until all large file transfers are finished. Consequently the total simulation time depends on the end of large flows transmission.

Table 3 shows the simulation parameters used for size-based fairness evaluation under different queue management policies. We simulated this scenario for different number of nodes starting from 6 nodes to 46 nodes to make different workloads and of course different number of short-lived and long-lived flows in system. Node numbers are chosen to be even so that they can be halved to short-lived and long-lived flow generators equally.

4.2. Direction-Based Fairness Evaluation. Ten mobile stations are placed within 5 m of the AP. Half of mobile stations are TCP senders. They upload data to the server using fixed-size data packets (1460 bytes). The other wireless stations are TCP clients, downloading data from the server using the same TCP data packet size as the upstream flows. We run each simulation for duration of 300 seconds simulation time to determine steady state throughput. Since AP buffer size is a dominant factor determining the degree of direction-based fairness in WLANs we repeated our simulation under different AP buffer sizes starting from 5 packets to 200 packets.

4.3. Assumptions. The following assumptions have been made in evaluation of our proposed TLAS-SAF scheme using mentioned simulation scenarios.

(i) All mobile nodes are fixed in their positions under our simulation scenarios. We are not concerned about mobility models since this is out of paper scope.

(ii) All radios of mobile stations are turned to 802.11b channel.

(iii) No interference or noise exists over wireless channel.

(iv) No propagation error model is assumed.

(v) No link/channel failure is assumed.

(vi) All wireless stations are in range of AP and can communicate data to and from AP.

4.4. Performance Metrics. In this paper, three performance metrics have been used to evaluate the performance of our proposed TLAS-SAF queue management policy compared to the LAS, TLAS, and conventional DropTail/FIFO. These are: Mean transfer time, Coefficient of variation of transfer time, and Jain's Fairness index [29]. The first two metrics have been employed to evaluate the size-based fairness and the last one is used to evaluate the direction-based fairness. Jain's fairness index is calculated by

$$\mathrm{FI} = \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n \sum_{i=1}^{n} x_i^2} \tag{9}$$

which rates the fairness of a set of flows when there are n flows and x_i is the throughput for the ith flow. The results

Table 3: Simulation parameters for size-based fairness evaluation.

Simulation time	End of large files transmission
Number of nodes	6–46 nodes
Default AP buffer size	50 packets
Nodes generating short-lived flows	Half of total
Nodes generating long-lived flows	Half of total
Number of large flows per station	10
Number of short flows per station	1 (repeating)
Large flow size	1000 packets
Short flow size	10 packets
Large flow arrival time	Sequentially
Short flow distribution	Exponential ON-OFF
Short flow burst time	250, 500, 1000 ms
Short flow idle time	250, 500, 1000 ms

Table 4: Simulation parameters for direction-based fairness evaluation.

Simulation time	300 seconds
Number of nodes	10 nodes
AP buffer size	5–200 packets
Nodes generating upstream flows	Half of total
Nodes generating downstream flows	Half of total
Number of flows per station	1 (continuously)

range from $1/n$ as the worst case and 1 as the best case when all flows receive the same allocation of bandwidth.

5. Results and Discussions

In this section, we present the results of simulations that are processed by ns-2 according to the simulation parameters and setup defined previously. In addition, we discuss and analyze the graphs generated based on simulation outputs.

5.1. Size-Based Fairness Evaluation. In first set of our simulations, we evaluated the Coefficient of Variation (CoV) of transfer time for large files and small files when conventional DropTail/FIFO queue management policy applied at AP buffer to demonstrate the size-based unfairness phenomenon. Figure 4 shows the CoV of transfer time for small files and large files when the number of nodes is varied from 6 to 46 and DropTail/FIFO applied at AP downstream queue. From this figure, it is obvious that presence of long-lived flows led to the higher variation of transfer time of small files while existence of short-lived flows had no specific impact over large file transfers.

The situation gets even worse for short-lived flows when number of wireless nodes increases. We can deduce that under DropTail/FIFO queue management policy in a mixed traffic scenario, short-lived flows suffer from higher variation of transfer time compared to the long-lived flows as previously justified. To solve this problem Least Attained Service scheduling is introduced which gives priority to the flows received the least service so far. The first packet to be served under LAS at AP downstream queue is a packet that its flow has received the minimum service among all other active flows in network.

However, LAS can improve the small file transfer time CoV to a high extent, it leads to the starvation of long-lived flows (and consequently large-file transfers). This drawback of LAS is alleviated by introducing another variant of LAS which is called Threshold-based LAS (TLAS) which gives service priority to network flows until a certain *Threshold* and it applies DropTail/FIFO over the rest of flows that have passed the threshold.

Figure 5 shows the CoV of small files and large files transfer time under our proposed TLAS-SAF and compares its performance with TLAS, LAS and, previously studied DropTail/FIFO policy. Figure 5(a) clarifies that LAS can alleviate the unfairness phenomenon against short-lived flows. While small files CoV of transfer time under DropTail/FIFO increased proportionally to the number of wireless nodes and reached to 2.96 for 46 nodes, LAS was able to keep this to a certain level of below 0.5 for all numbers of wireless stations. In contrast, LAS deteriorated large file transfers since it increased large files CoV of transfer time proportionally to the number of wireless nodes resulting in CoV of 1.7 for 46 nodes.

It is shown that while TLAS is able to keep small files CoV of transfer time at a constant level (between 0.6 to 1.2) for all numbers of wireless stations, it also achieves a good performance of large file transfers since it maintains a steady CoV level of below 0.5 for large file transfers. However TLAS can improve the CoV of small files transfer time when 46 wireless nodes are in WLAN from 2.96 as in DropTail/FIFO to 1.2, it compromises for almost 0.5 in CoV compare to LAS to alleviate the starvation phenomenon of long-lived flows. Since CoV of below one is assumed to be *low variation* the difference between small files CoVs of LAS and TLAS can be seen negligible.

Figure 5 also reveals that TLAS-SAF achieves comparable results to TLAS for both small and large file transfers. Similar to TLAS, TLAS-SAF is able to achieve a certain level of CoV of transfer time for small files (0.6 to 1.3) while not deteriorating large file transfers. The reason why we choose TLAS-SAF over TLAS is due to the fact that TLAS is unable to guarantee direction-based fairness while TLAS-SAF is designed to provide both types of fairness.

To investigate the performance of TLAS-SAF under different traffic patterns, we repeated our simulations for different inter-arrival times of short flows. Figures 6 and 7 show the CoV of transfer time for small and large files when inter-arrival times of short flows are set to 250 ms and 1 second, respectively. These figures indicate that similar to TLAS, TLAS-SAF is able to keep the CoV of transfer time for

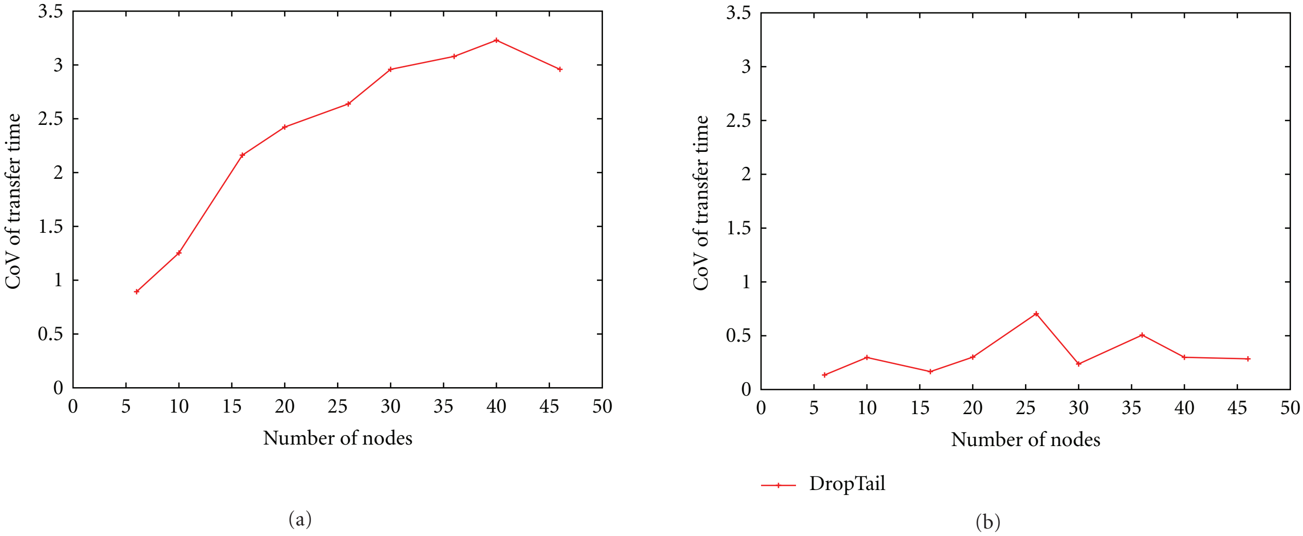


FIGURE 4: CoV of transfer time under DropTail/FIFO (short flows interarrival time = 500 ms): (a) small files; (b) large files.

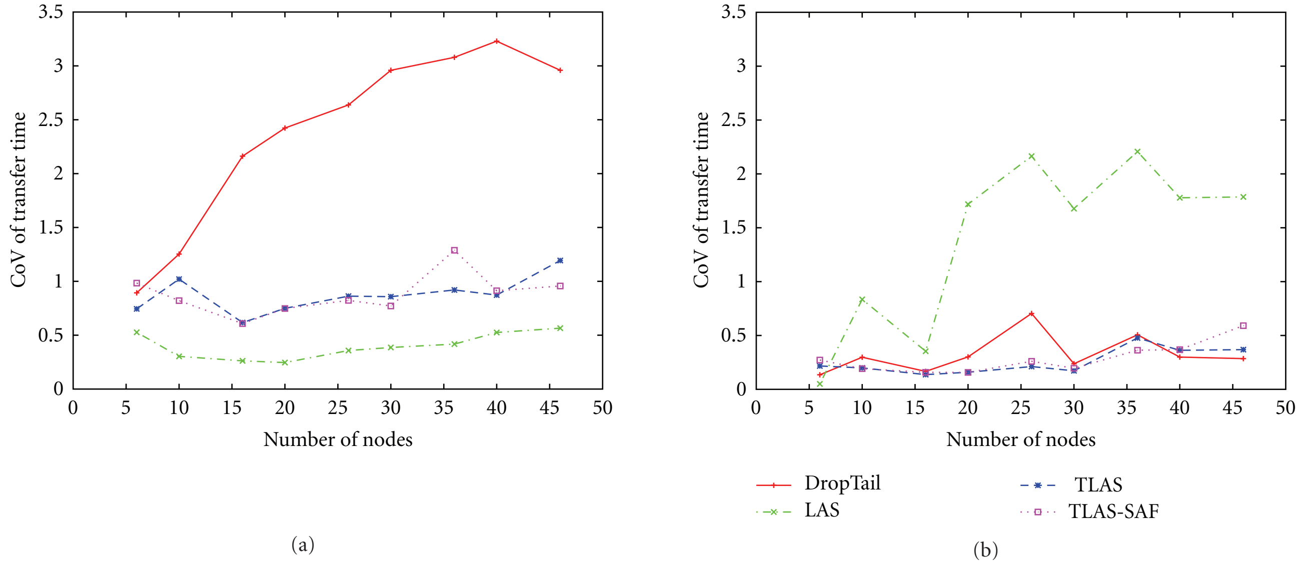


FIGURE 5: CoV of transfer time under TLAS-SAF and other studied policies (short flows interarrival time = 500 ms): (a) small files; (b) large files.

both short and large flows at certain level bound below one under both higher and lower traffic volume of short flows and different number of wireless nodes.

We considered an unexpected behavior in CoV of large files in LAS under inter-arrival times of 250 ms (Figure 6(b)) when their CoV of transfer time decreases for number of nodes higher than 30. We justify this behavior by knowing the fact that all large flows possibly get starved equally in LAS under high load of short flows since short flow packets occupy the buffer almost permanently. Therefore, with increase of large files transfer time to a very high number, their CoV tends to decrease. We further observe this issue when we study the mean transfer time of these flows.

Figures 8 and 9 demonstrate the mean transfer time of small and large file transfers, respectively, for TLAS-SAF and other queue management policies. While small files mean transfer time under DropTail/FIFO increases proportionally to the increase of number of nodes, LAS could reduce it to a maximum bound of below 0.5 second under different number of flows and inter-arrival times. But as we justified previously, large files mean transfer time get deteriorated compared to DropTail/FIFO for short inter-arrival times and hence under higher traffic (Figure 9(a)). It only becomes almost equal to DropTail/FIFO when we increased the short flow inter-arrival time to 1 second and therefore generated low-volume traffic (Figure 9(c)).

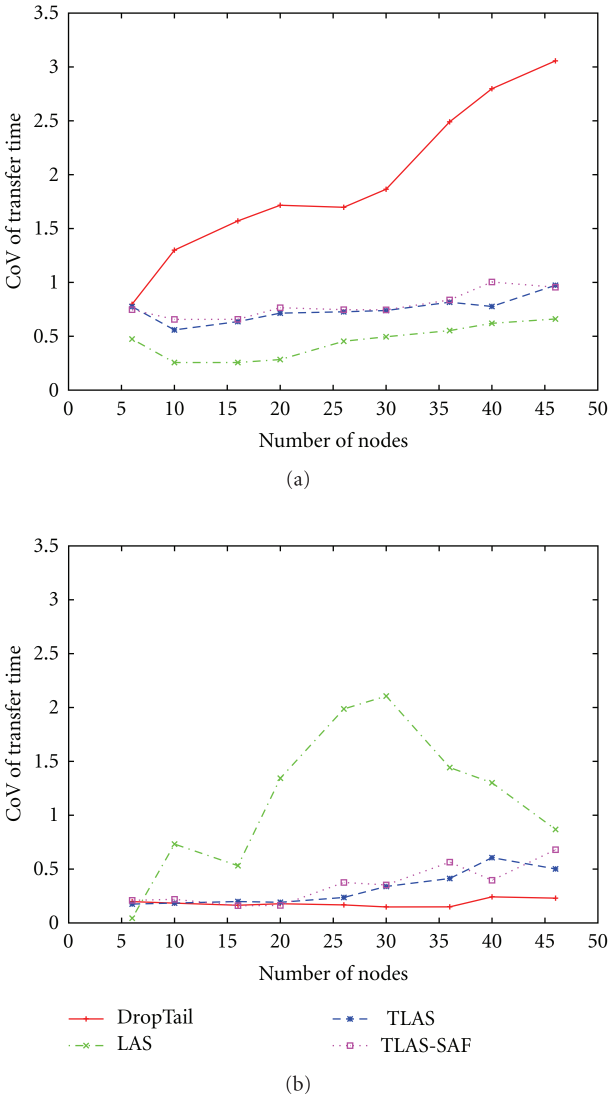


Figure 6: CoV of transfer time under TLAS-SAF and other studied policies (short flows inter-arrival time = 250 ms): (a) small files. (b) large files.

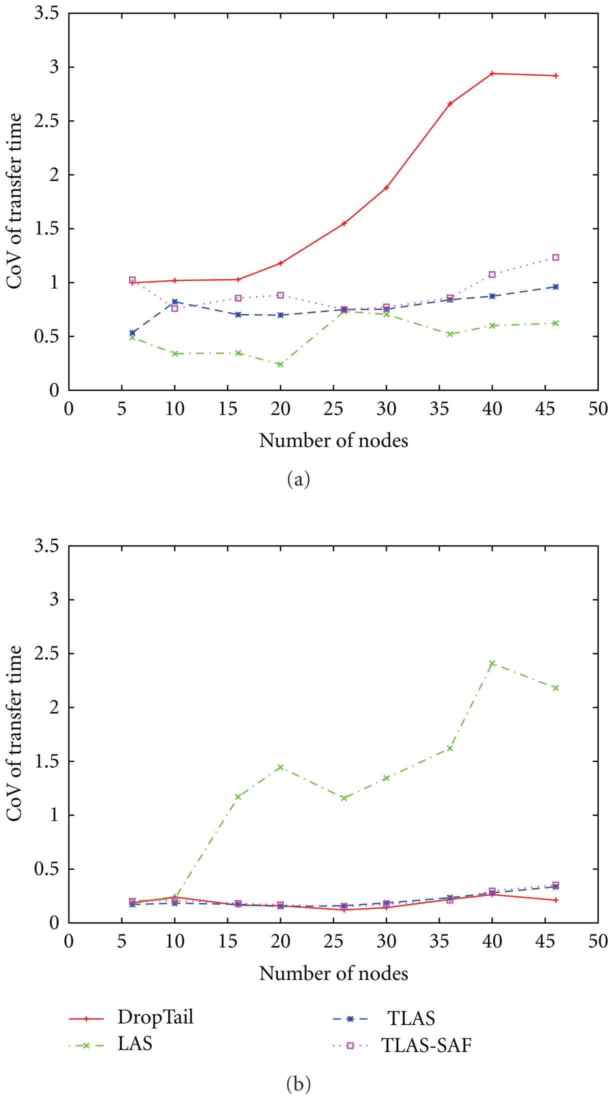


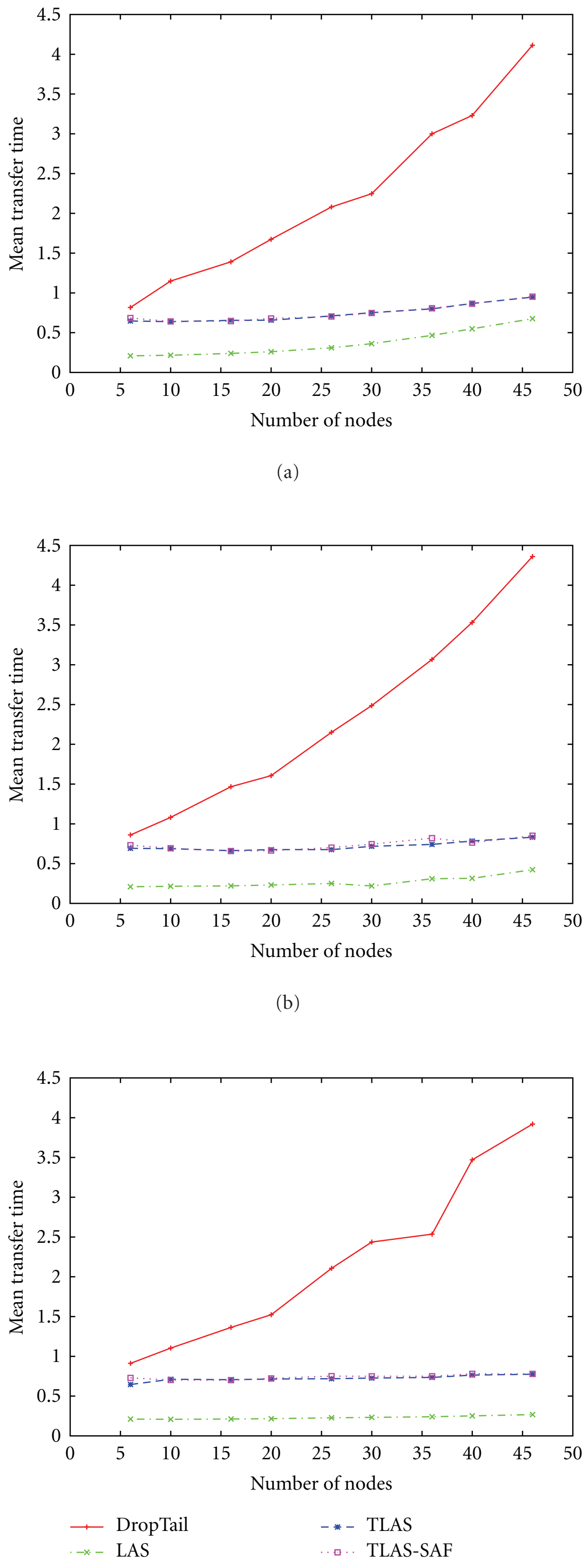
Figure 7: CoV of transfer time under TLAS-SAF and other studied policies (short flows inter-arrival time = 1 second): (a) small files. (b) large files.

All these issues justify our previous claim that LAS is unable to provide size-based fairness under high load. On the other hand, TLAS and TLAS-SAF could maintain a trade-off between DropTail/FIFO and LAS while compromising around 0.5 second for small files and reducing large files transfer times significantly compared to LAS. The close performance results of TLAS and TLAS-SAF proves our previous claim that TLAS-SAF inherits the characteristics of TLAS under high variance network flow sizes.

Regarding the previously mentioned anomaly of CoV of large flows under high traffic in LAS, the study of mean transfer time proves our justification as LAS mean transfer time increased sharply with increase of number of nodes due to the equal starvation of all long-lived flows (Figure 9(a)).

5.2. Direction-Based Fairness Evaluation. In second set of simulations, we calculated the Jain's Fairness Index for a scenario consisting of upstream and downstream flows sharing an AP downstream queue bottleneck. Since AP buffer size is a dominant factor determining the degree of fairness among these flows, we repeated this simulation for AP buffer size range of 5 to 200 packets. Figure 10(a) shows the fairness index of TCP flows when DropTail/FIFO queue management policy applied at AP buffer. It indicates that for small buffer sizes, presence of upstream flows leads to the unfair bandwidth allocation against downstream flows.

If AP buffer size increases, there will be more buffer space to accommodate downstream packets resulting in better fairness index. Fair bandwidth allocation happens when AP buffer size is 200 packets. The reason behind this issue is that in ns-2 TCP advertised window is set to 20 by default. This means at any given moment there would be maximum 20 data packets in transit from AP downstream queue for each individual flow.

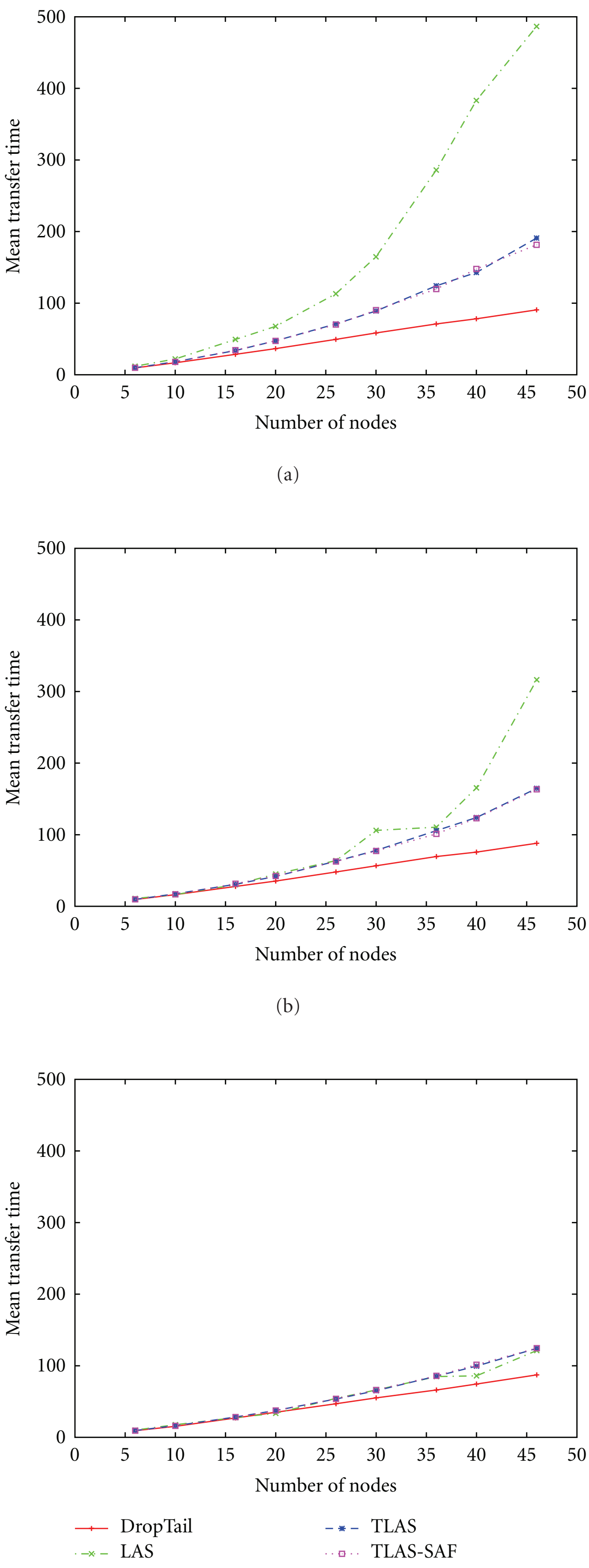


(a)

(b)

(c)

FIGURE 8: Mean transfer time of small files for short flow interarrival time: (a) 250 ms; (b) 500 ms; (c) 1 second.

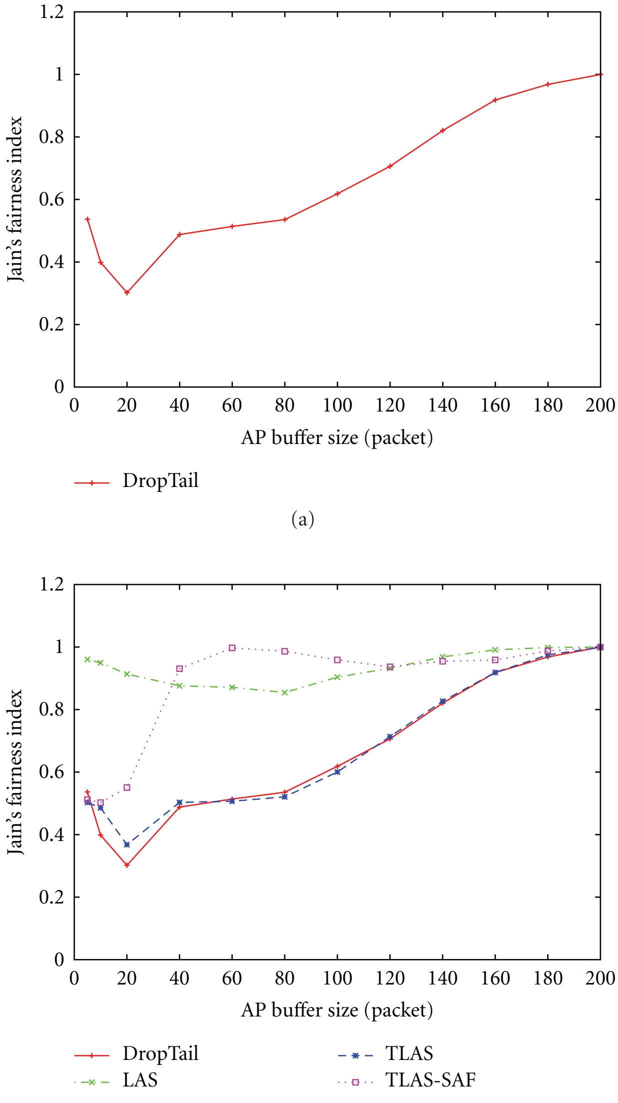


(a)

(b)

(c)

FIGURE 9: Mean transfer time of large files for short flow interarrival time: (a) 250 ms; (b) 500 ms; (c) 1 second.



(a)

(b)

Figure 10: Jain's fairness index: (a) DropTail/FIFO. (b) TLAS-SAF and other policies.

Since we have 5 stations each of which generating one downstream flow, the total TCP data packets in transit will be 100 packets. The other 5 stations (upstream flow generators) receive ACK packets via AP. In ns-2, the number of segments per ACK is set to one by default (however, in reality it is set to two segments per ACK). This means 100 ACK packets need to be transited via AP downstream queue giving the total 200 packets based on

$$B \geq (\partial MW + NW) = 5(20) + 5(20) = 200, \qquad (10)$$

where ∂ is number of ACK packet per data packets in TCP and W is the TCP-advertised receiver window for all flows and N and M refers to the number of downstream and upstream flows, respectively, as specified in [5].

In fact, AP buffer sizes are designed to be small. A large buffer size results in longer queuing delay as well as complex architecture and implementation issues and higher cost. For high-speed WLANs the situation gets even worse since the total bandwidth used by a wireless router results in very large buffer spaces. Thus, LAS policy is introduced to solve the unfairness for small buffer sizes. Figure 10(b) shows Jain's fairness index of network flows similar to the previous simulation scenario when proposed TLAS-SAF, LAS, and TLAS employed in AP buffer.

It indicates that LAS achieves a good level of fairness (Jain's index between 0.85 and 1.0) while TLAS behaves as worse as DropTail/FIFO for the whole range of AP buffer sizes. As we explained, employing LAS cannot guarantee size-based fairness. On the other hand, TLAS is also unable to provide direction-based fairness. Based on Figure 10(b) we can deduce that TLAS-SAF was able to provide fair bandwidth allocation among downstream and upstream flows for AP buffer sizes above 40 packets. It achieved a fairness index in range of between 0.93 and 1.0 for these AP buffer sizes outperforming LAS in term of overall fairness. However, fairness index declined to 0.51 when AP buffer size decreased to 5 packets only.

Since there are 10 flows in this scenario, the minimum required buffer size to guarantee a minimum congestion window of four packets for each flow will be 40. If buffer size is smaller than 40 at least one of the flows will always have a congestion window less than four resulting in permanent ACK dropping based on TLAS-SAF algorithm. Consequent ACK drops from the queue will eventually lead to the retransmission of packets or RTO and hence results in lower fairness index. By the way, most AP buffer sizes in current trend of market are larger than 50 packets and this issue does not form a drawback to employment of proposed TLAS-SAF in AP buffers.

6. Conclusion and Future Works

Several solutions have been proposed to solve the two main types of unfairness caused by cross-layer interaction of DCF mechanism of IEEE 802.11 MAC protocol and TCP transport protocol. These two types of unfairness are in benefit of long-lived versus short-lived and upstream versus downstream flows. The majority of Internet traffics are small file web transfers which are downloaded from the Internet to the wireless clients via AP bottleneck. Proposed solutions in the literature are mostly based on size-based scheduling policies which give service priority to small-size flows. However, these solutions can solve any of discussed unfairness issues, they are still unable to guarantee both types of required fairness.

To solve this issue, we proposed TLAS-SAF queue management policy to be applied at AP downstream queues. Simulations results show that TLAS-SAF is able to provide better service for short-lived flows while not deteriorating long-lived flows. These achievements in service improvement are in terms of mean and variation of transfer time and queuing delay experienced. Based on simulation results, we observed that TLAS-SAF was able to provide a good degree of fairness among upstream and downstream flows achieving the closest fairness index to one for most of the AP buffer sizes compared to other studied queue management policies.

Consequently we can conclude that TLAS-SAF can be taken into account as a unique solution for both size-based and direction-based fairness issues in IEEE 802.11 WLANs while other proposed solutions in the literature can only perform well in one of these aspects.

Future trend of technology in 802.11 WLANs provide more bandwidth and better quality of service to end users. However, we focused on 802.11b only, we believe that more research activity should be taken to focus on fairness issues in high-speed 802.11n and QoS-oriented 802.11e WLANs. In addition to proposed queue management policies at access point, adaptive dynamic buffer sizing can be seen as an interesting solution to prevent buffer spaces to become bigger in high-speed WLANs. In the future, we will try to focus on interaction of newer 802.11 MAC protocols and TCP variants to evaluate the performance of our proposed model. In addition to simulations, test-bed experiment results must be taken into consideration to prove the performance of TLAS-SAF in real networks.

Acknowledgment

The authors would like to thank the Ministry of Higher Education of Malaysia for its financial support of this publication under the Fundamental Research Grant (FRGS) 02-01-07321FR.

References

[1] IEEE Computer Society LAN MAN Standards Committee, "IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," IEEE Std 802.11-1997, The Institute of Electrical and Electronics Engineers, 1997.

[2] R. Bruno, M. Conti, and E. Gregori, "Throughput evaluation and enhancement of TCP clients in Wi-Fi hot spots," in *Proceedings of the 1st Working Conference on Wireless On-demand Network Systems (WONS '04)*, pp. 73–86, Madonna di Campiglio, Italy, 2004.

[3] Q. Wu, M. Gong, and C. Williamson, "TCP fairness issues in IEEE 802.11 wireless LANs," *Computer Communications*, vol. 31, no. 10, pp. 2150–2161, 2008.

[4] M. Bottigliengo, C. Casetti, C.-F. Chiasserini, and M. Meo, "Short-term fairness for TCP flows in 802.11b WLANs," in *Proceddings of the 23rd IEEE Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '04)*, pp. 1383–1392, Hong Kong, March 2004.

[5] S. Pilosof, R. Ramjee, D. Raz, Y. Shavitt, and P. Sinha, "Understanding TCP fairness over wireless LAN," in *Proceedings of the 22nd Annual Joint Conference on the IEEE Computer and Communications Societies*, pp. 863–872, San Francisco, Calif, USA, April 2003.

[6] D. J. Leith and P. Clifford, "Using the 802.11e EDCF to achieve TCP upload fairness over WLAN links," in *Proceedings of the 3rd International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt '05)*, pp. 109–118, Trentino, Italy, April 2005.

[7] X. Lin, X. Chang, and J. Muppala, "VQ-RED: an efficient virtual queue management approach to improve fairness in infrastructure WLAN," in *Proceedings of IEEE Workshop on Wireless Local Networks (WLN '05)*, pp. 632–638, Sydney, Australia, 2005.

[8] A. Ng, D. Malone, and D. Leith, "Experimental evaluation of TCP performance and fairness in an 802.11e test-bed," in *Proceedings of the ACM SIGCOMM Workshop on Experimental Approaches to Wireless Network Design and Analysis*, pp. 17–22, Philadelphia, Pa, USA, 2005.

[9] L. Guo and I. Matta, "The war between mice and elephants," Tech. Rep. BU-CS-2001-005, Computer Science Department, Boston University, 2001.

[10] S. Ebrahimi-Taghizadeh, A. Helmy, and S. Gupta, "TCP vs. TCP: a systematic study of adverse impact of short-lived TCP flows on long-lived TCP flows," in *Proceddings of the IEEE Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '05)*, vol. 2, pp. 926–937, Miami, Fla, USA, March 2005.

[11] K. Tokuda, G. Hasegawa, and M. Murata, "Analysis and improvement of the fairness between long-lived and short-lived TCP connections," in *Proceedings of the IEEE Workshop on Protocols for High-Speed Networks (PfHSN '02)*, pp. 151–158, 2001.

[12] M. Harchol-Balter, B. Schroeder, N. Bansal, and M. Agrawal, "Size-based scheduling to improve web performance," *ACM Transactions on Computer Systems*, vol. 21, no. 2, pp. 207–233, 2003.

[13] I. Rai, G. Urvoy-Keller, M. Vernon, and E. Biersack, "Performance models for LAS-based scheduling disciplines in a packet switched network," in *Proceedings of the ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, pp. 106–117, New York, NY, USA, 2004.

[14] I. Matta and L. Guo, "Differentiated predictive fair service for TCP flows," Tech. Rep. BU-CS-2000-012, Computer Science Department, Boston University, 2000.

[15] S. Pollin, A. Motamedi, A. Bahai, F. Catthoor, and L. D. Van Perre, "Delay improvement of IEEE 802.11 distributed coordination function using size-based scheduling," in *Proceedings of the IEEE International Conference on Communications (ICC '05)*, vol. 5, pp. 3484–3488, May 2005.

[16] M. Gong, Q. Wu, and C. Williamson, "Queue management strategies to improve TCP fairness in IEEE 802.11 wireless LANs," in *Proceedings of the 4th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*, pp. 1–8, 2006.

[17] V. Paxson and S. Floyd, "Wide area traffic: the failure of Poisson modeling," *IEEE/ACM Transactions on Networking*, vol. 3, no. 3, pp. 226–244, 1995.

[18] K. Claffy, G. Miller, and K. Thompson, "The nature of the beast: recent traffic measurements from an Internet backbone," in *Proceedings of INET*, 1998.

[19] F. Liang, X. Wang, L. Xu, and J. Fan, "On AP buffer effect upon TCP fairness over WLAN," in *Proceedings of the IEE Mobility Conference*, p. 18, November 2005.

[20] X. Wang and F. Liang, "TFBR: a new queue management guaranteeing TCP fairness based on TCP advertised receiving windows over WLAN," in *Proceedings of the 2nd International Conference on Mobile Technology, Applications and Systems*, pp. 1–7, 2005.

[21] T. Li and D. L. Leith, "Buffer sizing for TCP flows in 802.11e WLANs," *IEEE Communications Letters*, vol. 12, no. 3, pp. 216–218, 2008.

[22] I. A. Rai, E. W. Biersack, and G. Urvoy-Keller, "Size-based scheduling to improve the performance of short TCP flows," *IEEE Network*, vol. 19, no. 1, pp. 12–17, 2005.

[23] I. A. Rai, G. Urvoy-Keller, and E. W. Biersack, "Analysis of LAS scheduling for job size distributions with high variance,"

Performance Evaluation Review, vol. 31, no. 1, pp. 218–228, 2003.

[24] G. Urvoy-Keller and A.-L. Beylot, "Improving flow level fairness and interactivity in WLANs using size-based scheduling policies," in *Proceedings of the 11th International Conference on Modeling, Analysis, and Simulation of Wireless and Mobile Systems (MSWiM '08)*, pp. 333–340, ACM, New York, NY, USA, October 2008.

[25] C. Williamson and Q. Wu, "A case for context-aware TCP/IP," *ACM Performance Evaluation Review*, vol. 29, no. 4, pp. 11–23, 2002.

[26] I. Rai, *QOS support in edge routers*, Ph.D. thesis, Télécom Paris, Institut Eurecom, Paris, France, September 2004.

[27] E. W. Biersack, B. Schroeder, and G. Urvoy-Keller, "Scheduling in practice," *Performance Evaluation Review*, vol. 34, no. 4, pp. 21–28, 2007.

[28] J. Fall and K. Varadhan, "The NS Manual," 2006, http://www.isi.edu/nsnam/ns/ns-documentation.html.

[29] R. Jain, D. Chio, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," Tech. Rep. DEC RR TR-30I, DEC Research, 1984.