Selecting a Preferable Access Point with More Available Bandwidth

Shibo Xu, Fengyuan Ren, Yinsheng Xu, Chuang Lin, Min Yao

Tsinghua National Laboratory for Information Science and Technology,

Dept. of Computer Science and Technology, Tsinghua University, Beijing, 100084, China

Abstract—It is a common problem for Wireless LAN (WLAN) users when they face more than one available Access Point (AP): which one may serve them better? In the real world, most WLAN devices and users select the AP by Received Signal Strength (RSS), which doesn't consider the load of APs. Therefore, the RSS-based scheme may result in load imbalance and utilization degradation. This paper proposes a simple scheme to inform WLAN users the available bandwidth when selecting a specific AP. Unlike the previous work, the new scheme concerns about the aggregated traffic patterns in AP, which affects the amount of the available bandwidth. Considering the number of active stations and influence of collision, the available bandwidth is calculated and regarded as a selection metric. Common users can acquire this metric through Service Set IDentifier (SSID) without any modifications on their devices, and users who accept modifications can select the preferable AP dynamically and automatically. This scheme is very simple to be implemented and deployed with negligible overhead to network. We conduct simulations to verify this scheme, which can improve by 200% the user's throughput in some situations.

Index Terms—IEEE 802.11 WLAN, Access Point Selection, Available Bandwidth, Active Station.

I. INTRODUCTION

The convenience, flexibility and lowcost of WLAN advantage its large-scale deployment in airport, marketplace, campus and office. There are two operation modes in WLANs, the basic mode and the optional mode. The former is compulsory, which uses two-way handshake (DATA-ACK). The latter uses Request-To-Send (RTS) and Clear-To-Send (CTS) in a four-way handshake (RTS-CTS-DATA-ACK). Since turning on RTS/CTS may incur a significant overhead, in most cases, WLANs operate in the basic mode. In this paper, our investigation is accomplished in the basic mode. In WLANs, there are two types of devices, user devices (such as laptop, tablet and smartphone) and APs. Hereinafter, the user devices are generally called stations. The stations connect to networks through APs, and APs relay data just like a router.

In WLANs, a station acquires information about APs through two ways, active and passive. In the active way, the station broadcasts probe request frames, and APs send their configuration back to the station in probe response frames. In the passive way, AP broadcasts beacon frames containing configuration information periodically. Each WLAN is identified by a SSID, which is an ASCII string with maximum length of 32 bytes. SSID is used when a station associates with an AP. If the SSID provided by the station is different from that in the AP, the AP will reject the station.
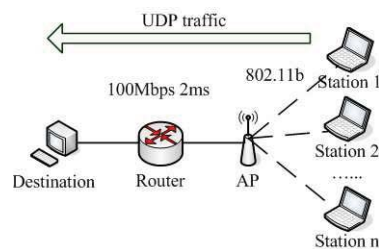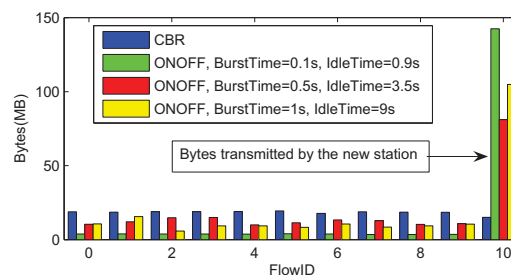

Fig. 1. Simulation Topology

Fig. 2. Throughput under Different Traffic Pattern

Before data transmission, a station needs to associate with only one AP. Generally, a station may scan several APs which provide different potential bandwidth, so the selection of AP is a critical problem. The default method for AP selection is only based on RSS, and stations will select the AP with strongest RSS, while ignoring the load status of APs. This onefold criterion readily results in imbalanced distribution of traffic among APs, lower utilization of WLANs and worse user experience. Although RSS is an important indicator of channel quality between station and AP, it can not indicate the available bandwidth that the AP can provide. In this paper, we define the available bandwidth of an AP as the maximum throughput achievable to an incoming station when associating with the AP.

Some existing methods detect AP's load by its throughput, response delay or the number of associated stations. But they need modifications on APs and stations, or incur extra overhead to the network. Furthermore, they neglect the traffic patterns in APs. Evidently, the available bandwidth of an AP is related to its capability and load. Nevertheless, based on numerous experimental observations, we find that the available

bandwidth which can be captured by an incoming station is greatly affected by the aggregated traffic pattern in the AP. For example, we conduct some simulations on the NS2 platform, and the topology is shown in Fig. 1. We activate 10 stations four times in the same topology. In each time, one flow with equivalent rate (1Mbps) is activated on each station. Those 10 stations run for 5 minutes as background traffic, while a new greedy station with a CBR flow is activated to compete for bandwidth. The new station is activated at 1 minute and lasts for 4 minutes. The successfully transmitted bytes of each station are shown in Fig. 2, where the flow with ID 10 is the new one. It is clear that the new flow can capture more bandwidth when the background are ONOFF traffics, and CBR flow is aggressive to ONOFF flow. The result also indicates the fairness of homogeneous flows. Although the APs are equally overloaded, the throughput of the new station significantly depends on the aggregated traffic pattern.

Based on the above observations, the main contributions of this paper are threefold:

- We consider the available bandwidth as a new selection metric, which is calculated by monitoring the aggregated traffic on AP without any probe overhead. Some methods are proposed to estimate the available bandwidth by taking the traffic patterns, the number of active stations and collisions into account.
- A new mechanism is designed to inform the selection metric to stations and users. According to this metric, stations can select or switch to the AP with more available bandwidth without previous association with the APs.
- A simple and effective scheme for AP selection is proposed, which supports two operation modes, single-side and dual-side. The single-side mode doesn't need any modifications on stations, and users may select the preferable AP according to its SSID manually. In the dual-side mode, some modifications (e.g., installing a software) are required on stations, which can select or switch to a better AP automatically and dynamically.

The rest of this paper is organized as follows. In Section II, we make a summary of previous work on AP selection. The new scheme and related computing methods are introduced in Section III. The implementation issues are discussed in Section IV. In Section V, the comparative simulations are conducted to validate the efficiency of the proposed scheme. Finally, the paper is concluded and some discussions are presented in Section VI.

II. RELATED WORK

A lot of work has been done to address the issue of AP selection, which can be classified into two types according to different rules.

The number of used APs: single-AP or multi-AP. Most earlier work provided single-AP solutions, which use one AP as router in a period, such as [1] [2] [3] [4] [5]. Some recent research introduces multi-AP solutions to employ more than one AP in a short time, such as [6] [7] [8] [9]. If the bottleneck is the backhaul link, multi-AP solutions can provide more aggregated bandwidth than single-AP solutions. Provided that there are some stationary stations and APs, single-AP solutions can achieve similar performance as multi-AP solutions with less cost if properly configured. If the bottleneck is wireless link, the single-AP solutions are superior to multi-AP solutions.

The way of measurement: monitoring or probing. The monitoring method just inspects the relevant data without sending any segments, which does not introduce overhead to network, such as [1] [2] [4]. The probing method needs to send some requests to middleware or destination, and can get more information by analyzing the responds, which is costful, such as [3] [5].

The location of modifications to implement: single-side or multi-side. In general, some modifications are necessary to implement and deploy solutions. The simple way is single-side, which just needs some modifications on stations or APs, such as [1] [5]. For multi-side, the modifications are needed on both stations and APs. Moreover, it needs to develop and deploy new components in the network, such as [2] [3] [4].

Because of insufficiency of the RSS-based scheme, more useful metrics are considered in literatures, such as delay [1], the number of associated stations [2], link quality [2], path performance [3], effect on itself and other stations when selecting an AP [4]. The potential bandwidth between station and AP is regarded as a metric in [5], which is estimated through the delay of beacon frames. However, it leaves the number of stations out of account and does not analyze deeply the impact of collisions. Overall, all those schemes haven't investigated the effects of traffic patterns, and they need modifications on stations or incur probe overhead to networks.

Considering such a common scenario, a lot of users attempt to select an AP in a building, and the wireless link is bottleneck. Since user devices are usually self-configured, it is better to provide an optional mode for AP selection to common users. Therefore, the ideal solution is single-AP, monitoring and single-side (AP-side), which simplifies the implementation and deployment. The proposed scheme fulfills these requirements, which monitors the aggregated traffic on AP without sending any extra segments, and stations may select a preferable AP without any modifications on their devices.

III. PROPOSED SCHEME

A. Basic Idea

Generally, in WLANs, a station with empty queue is silent. Otherwise it is active. The MAC layer of IEEE 802.11 can distribute transmission opportunities among active stations fairly. If no station is active, the incoming station will make full use of the AP. Otherwise it can only get it's fair share. Therefore, the number of active stations is not only an effective indicator of AP's load, but also an important metric of potential transmission opportunity. In the previous simulations, the average number of active stations counted within different timespan is shown in Fig. 3. Obviously, the available bandwidth is inversely proportional to the number
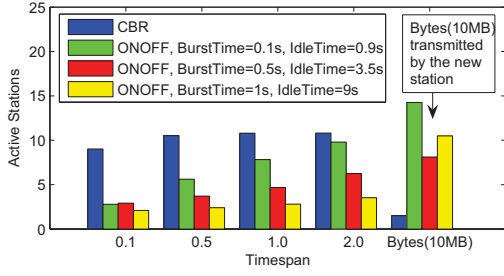
Fig. 3. The Average Number of Active Stations within Different Timespan
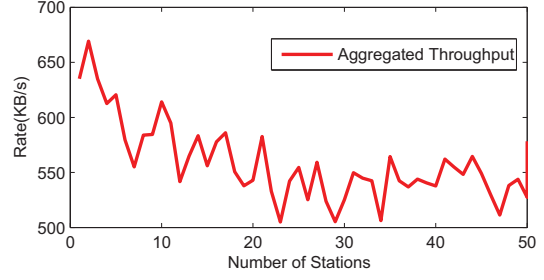

Fig. 4. Aggregated Throughput vs Number of Stations.

of active stations in spite of some deviation. The number of active stations is proportional to the length of timespan.

Let AC be the available capacity of the wireless channel, i.e., the maximum aggregated bandwidth for all stations, n be the number of active stations in a short timespan, and μ be the utilization ratio of the available capacity. Since IEEE 802.11 DCF distributes channel fairly among competitors, if the AP is overloaded ($\mu = 1$), each station achieves its fair share. When a new greedy station associates with the AP, the available bandwidth for the station is $\frac{AC}{n+1}$. If the AP is underloaded ($\mu < 1$), the station will obtain the unused bandwidth, i.e., $(1-\mu)AC$. Furthermore, the station will obtain additional bandwidth $\frac{\mu AC}{n+1}$ by competing with the existing n stations. Namely, the available bandwidth AB for the new station is

$$AB = \frac{\mu AC}{n+1} + (1-\mu)AC, 0 \le \mu \le 1 \quad (1)$$

AB is a proper metric for AP selection, and the incoming station may achieve better performance by selecting the AP with higher AB. The basic idea of the proposed scheme is to calculate this metric (AB) on AP and advertise it to stations, then stations may make proper choices based on this metric. In the following subsections, methods for calculating AC and estimating n will be discussed.

B. Calculation of Available Capacity

Although the maximum transmission rate of IEEE 802.11 is very high, for example 11/54/150Mbps for 802.11b/g/n respectively, the available capacity is limited. The Theoretical Maximum Throughput (TMT) of IEEE 802.11 is estimated in [10], which is given in Equation (2). For the basic mode of 802.11b, $a = 0.72727$ and $b = 890.73$, s is size of MAC Service Data Unit (MSDU), if the MSDU is 1500 bytes, TMT will only be 739KB/s. Note that, the TMT is theoretical maximum throughput in MAC layer, and the overhead due to headers of interlayers should be gotten rid of, thus TMT for applications is further less.

$$\begin{aligned} TMT(s) &= \frac{MSDU\ size}{Delay\ per\ MSDU} \\ &= \frac{8s}{T_{DIFS} + T_{SIFS} + T_{BO} + T_{ACK} + T_{DATA}} \\ &= \frac{8s}{as+b} \times 10^6\ (bps) \end{aligned} \quad (2)$$

This theoretical analysis neglects collisions, but they indeed impact the available bandwidth. As well known, the collision probability is related to the number of stations, and the available bandwidth is related to collision probability. As shown in Fig. 4, along with growth of the number of greedy stations, the aggregated throughput of AP degrades obviously. Hence, the impact of collision on the available capacity should be taken into account.

Assume that each station has a collision probability p, W is the initial contention window, namely CW_{min}. The probability of successful transmission on the first attempt is $1 - p$, and it's $p^i(1-p)$ on the $(i+1)th$ attempt. Based on Binary Exponential Backoff (BEB) of DCF, the average backoff window W_{ave} is $\frac{W-1}{2}$ on the first attempt, and it's $\frac{2^i W - 1}{2}$ on the $(i+1)th$ attempt. Attempting will go on to M times until contention window reaches CW_{max}, so the overall average backoff window is

$$W_{ave} = \sum_{i=0}^{M-1} \frac{(1-p)p^i(2^i W - 1)}{2} \quad (3)$$

Thus, $\frac{1}{W_{ave}}$ is the probability for a station to transmit, and the transmission will succeed if no other stations transmit. If the number of active stations is n, the probability of successful transmission is $(1 - \frac{1}{W_{ave}})^{n-1}$. Then the collision probability is

$$p = 1 - (1 - \frac{1}{W_{ave}})^{n-1} \quad (4)$$

From Equation (3) and (4), it is obvious that p is a function of n, which can be approximated as

$$p = f(n) \approx 1 - \alpha^{n-1},\ 0 \le \alpha \le 1 \quad (5)$$

Then, the available capacity AC should be

$$AC = TMT\alpha^{n-1} \quad (6)$$

where α can be regarded as a modifying factor for collision.

C. Estimation of the number of Active Stations

In order to estimate n, a proper time window must be determined. If the time window is too large, it will impact the accuracy and timeliness of n. Besides, fairness is another precondition for precise estimation of n. Therefore, the time window should be as short as possible under the precondition
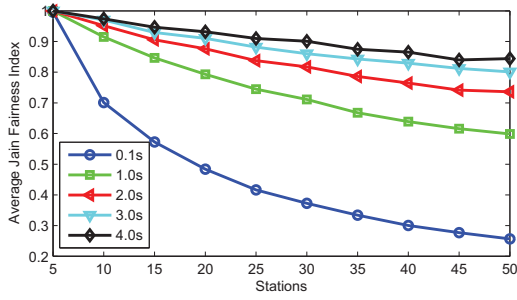
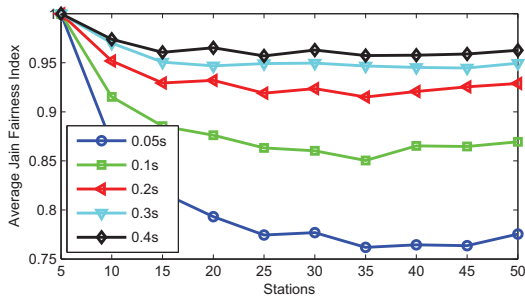Fig. 5. Average Jain Fairness Index for Different Timespan and Stations.

Fig. 6. Average Jain Fairness Index of Sliding Window with Different T_{unit}

Algorithm 1: calculating the number of active stations in a round.
1, set all items in Sta_Frames zero;
2, $n = 0$ and $T_{wnd} = N_{init} * T_{unit}$;
3, set timer with time length T_{wnd};
4, if a data frame is received or sent, get the index i of the station in Sta_Frames according to its address in frame;
5, if $Sta\_Frames[i] = 0$, $n = n + 1$;
6, $Sta\_Frames[i] = Sta\_Frames[i] + 1$;
7, if timer expire, set all items in Sta_Frames zero, else goto line 4;
8, $T_{wnd} = n * T_{unit}$, $n = 0$ and goto line 3.

and the number of active stations will not change greatly in a short time. Consequently, n in previous T_{wnd} can be used to calculate this T_{wnd}, and for the first time, a initial value N_{init} can be set to n.

IV. IMPLEMENTATION ISSUES

A. Calculation of n and μ

The method for calculation of n is given in Table I. A timer with duration T_{wnd} is set to trigger operation at the end of this sliding window. The array Sta_Frames is set to count the number of stations which have sent or received data frames. When the timer expires, n is obtained and the next T_{wnd} for the new timer is calculated. Since n obtained in every sliding window is dramatically variable, a T_{wnd} can be regarded as a round. The average value should be taken every some rounds for implementation, such as moving average.

μ is the utilization ratio of available capacity, i.e., $Aggregated\_Throughput/AC$, then it is readily obtained.

B. Notifying and Selecting

The available bandwidth AB mentioned above should be contained in probe response frames and beacon frames. We hope that our scheme introduces little overhead to WLANs and is suitable for unmodified stations. Hence, no additional fields are introduced. To an unmodified station, the only way to select a better AP relies on users' operation. In most cases, users can find some WLANs to use, and select one or switch to another one. The only information users can see is SSIDs of all WLANs. Thus, the proper way is to embed the metric in SSIDs. Common users who don't want any adjustment on their devices may operate as usual or select an AP in according to SSIDs manually. The software component for voluntary users collects the metrics from all APs, and selects or switches to the better one when the time is ripe.

The metric value is included in the SSID field, and another reserved bit is needed to indicate if the scheme is enabled. This paper makes use of the reserved bit (12) in the *Capability Information* field, as shown in Fig. 7. By default, a station will not change its AP unless the AP is unusable or the owner enforces a handoff. For convenience, the scheme makes the initial SSID as the common prefix for all dynamic SSIDs. For example, if the initial SSID of a WLAN is *Tsinghua*, the dynamic SSIDs will be *Tsinghua_XX*, where *XX* is our metric.

of fairness. The DCF in IEEE 802.11 is intend to distribute channel fairly to all stations, but in fact it is fair for long term, and unfair for short term [11]. Its fairness is related to the length of timespan and the number of active stations. Fig. 5 gives five curves of average Jain fairness index in NS2 simulations. Each of them is corresponding to a timespan, i.e., 0.1s, 1.0s, 2.0s, 3.0s and 4.0s for different number of greedy stations. As we can see that the fairness index is proportional to the length of timespan, and inversely proportional to the number of greedy stations. To select a proper time window, the number of active stations should be considered.

In order to keep fairness for variable numbers of stations, a sliding window T_{wnd} and a unit-window T_{unit} are introduced. The former is the duration of timespan, while the latter is the minimal computational unit, if the number of active stations is n, then the sliding window will be

$$T_{wnd} = n \times T_{unit} \qquad (7)$$

where n is the number of stations which have sent or received at least one data frame in the duration of T_{wnd}, i.e., n is

$$n = Number\_of\_Stations(T_{wnd}) \qquad (8)$$

To test this method, similar simulations are conducted, and the result is shown in Fig. 6. Each curve is corresponding to a T_{unit}, i.e., 0.05s, 0.1s, 0.2s, 0.3s and 0.4s. To keep average Jain fairness index above 0.90, it only need to set T_{unit} equal to 0.2s, 0.3s or 0.4s.

Combining Equation (7) and (8), T_{wnd} and n are interdependent. In a WLAN, the number of associated stations is limited,

B0	B1	B2	B3	B4	B5	B6	B7
ESS	IBSS	CF Pollable	CF-Poll Request	Privacy	Short Preamble	PBCC	Channel Agility

B8	B9	B10	B11	B12	B13	B14	B15
Spectrum Mgmt	QoS	Short Slot Time	APSD	Re-served	DSSS-OFDM	Delayed Block Ack	Imme-diate Block Ack

Fig. 7.   Capability Information Field

Then it is easy for the AP to identify whether a SSID comes from it or not. If the SSID provided by a station has the same prefix as AP, the AP will accept the station.

V. SIMULATION AND VALIDATION

To validate the proposed scheme, two groups of simulations have been conducted. The first group aims to check the fitness of the metric in a simple environment and a complicated environment. The second group aims to test the effectiveness of the scheme when AP is underloaded and overloaded. To avoid impacts caused by other factors, such as unfairness between uplink and downlink, unfairness of TCP flows, only UDP traffics are used and all traffics are from stations to AP. Packet size is used to calculate throughput and TMT, the mean value of which can be used in implementation and it is 1500 bytes in our simulations. There are three parameters, the unit-window T_{unit}, the initial number of active stations N_{init} and the modifying factor α. In most cases, stations join in a WLAN in a random order, and the initial number of active stations is limited, so $N_{init} = 2$. In all simulations, one flow is activated on each station, and the number of flows is equal to the number of stations.

In the first group of simulations, the metrics of AB are calculated under different background traffics, then a new greedy station is activated under same background traffic, finally the metrics and the throughput of the new station are compared. The simple background is incremental traffic. 50 flows are activated sequently by 2s interval, and each flow produces a CBR traffic of 1Mbps. The complicated background is random traffic. Where 2 CBR flows of 20Kbps run during the whole test to imitate multimedia applications, 48 ONOFF flows of 1Mbps are activated uniformly between 1s and 11s to imitate Web operations. The idle time and burst time of the ONOFF flows is exponentially distributed, whose average values are 25s and 1s respectively.

The results under simple and complicated background traffic are shown in Fig. 8 and Fig. 9, respectively. In the two figures, the green curve is the throughput of the new flow, the blue dashed curve is the metric with $T_{unit} = 0.1s$, and the red dashed curve is the metric with $T_{unit} = 0.2s$. As we can see, the metrics are close to the real throughput, and the large deviation appears just at the beginning. Since fairness is proportional to T_{unit}, the metric with $T_{unit} = 0.2s$ is closer to real result than the one with $T_{unit} = 0.1s$. As a whole, both of those metrics can give an appropriate indicator of the available bandwidth that the new flow can capture. By the way, in our
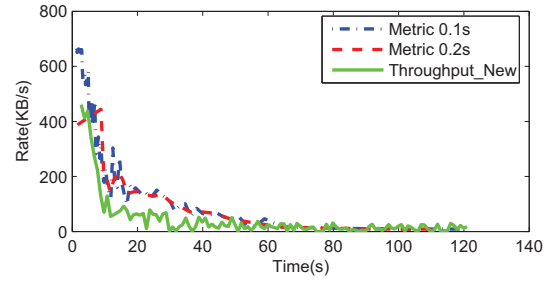

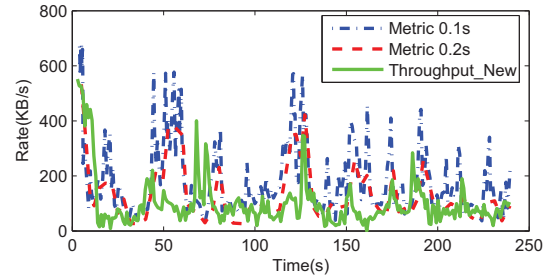Fig. 8.   Metric vs Throughput in Simple Background

Fig. 9.   Metric vs Throughput in Complicated Background

simulations, $\alpha = 0.99$, and we find the result is sensitive to α.

In the second group of simulations, assume that there are different number of users in two neighboring meeting rooms, as shown in Fig. 10. One AP is set in each room, and all users select the AP in this room. A new user with greedy traffic enters the room with more persons, and can select anyone of the two APs. To imitate RSS, if the user selects the AP in this room, the distance to the AP is about 5m, and if the user selects the AP in the next room, the distance is about 30m. Suppose that each background user produces an ONOFF traffic with 1Mpbs rate, average 1s burst time and average 2s idle time. For underloaded AP, 10 users in this room and 5 users in the next room. Throughput comparison is given in Fig. 11. For overloaded AP, 50 users in this room and 20 users in the next room. The comparison is given in Fig. 12. In both figures, the red curve is the throughput when the user selects the AP in the next room, and the blue dashed curve is the throughput when the user selects the AP in this room. Obviously, the user can capture more bandwidth when selecting the AP in the next room according to the proposed scheme.

VI. DISCUSSION AND CONCLUSION

Although mobility of stations affects the channel performance, it is negligible in most cases, especially in campus. According to conventional practice, other effect is also ignored, such as hidden nodes and exposed nodes. Some literatures have mentioned that TCP is seriously unfair in WLANs [12] [13] [14] [15], and our simulations have also verified it, but it is another problem. This paper is to advantage users to select
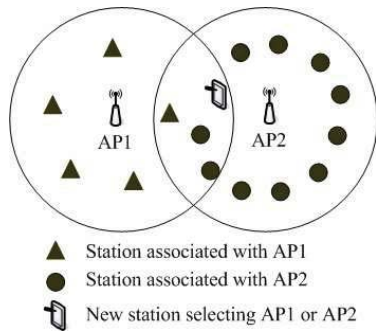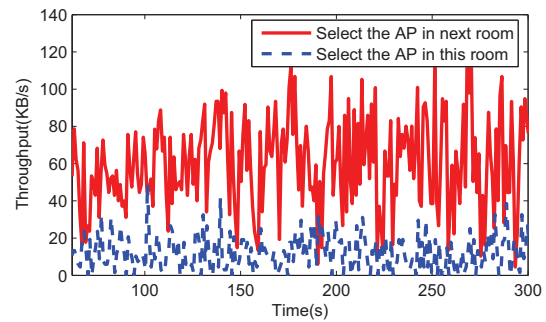
Fig. 10. Simulation Environment for AP Selection

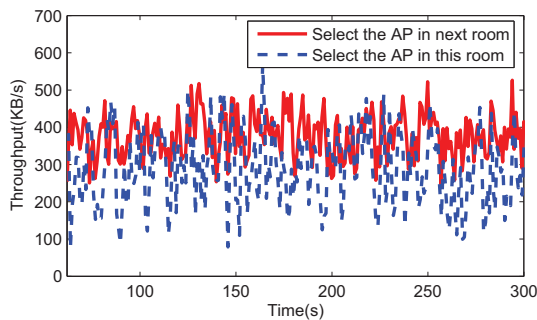Fig. 12. Throughput Comparison in Overloaded AP

Fig. 11. Throughput Comparison in Underloaded AP

an AP with more available bandwidth. AP selection for TCP based applications will be our future work.

Nowadays, most WLAN devices select AP by RSS, which is a good indicator of channel quality. However, this approach readily results in load imbalance and performance degradation. To address this issue, this paper proposes a simple scheme to inform users the available bandwidth they can obtain from APs. Users may make a better choice by combining the metric and RSS. The scheme is based on the fairness of DCF, and takes the number of active stations and impact of collision into account. As verified by simulations, the metric is an appropriate indicator of the available bandwidth. The scheme supports two operation modes, single-side and dual-side. Common users can get this metric through SSID, and voluntary users can select the preferable AP dynamically by a software.

VII. Acknowledgement

References

[1] V. A. Siris and D. Evaggelatou, "Access Point Selection for Improving Throughput Fairness in Wireless LANs," in *10th IFIP/IEEE International Symposium on Integrated Network Management, 2007. IM'07*, May 2007.

[2] S. Venkatesan and C. Manoharan, "Access Point Selection in WLAN using Optimal Association Control Algorithm," *European Journal of Scientific Research*, vol. Vol.60 No.2, April 2011.

[3] A. J. Nicholson, Y. Chawathe, M. Y. Chen, B. D. Noble, and D. Wetherall, "Improved Access Point Selection," in *MobiSys'06*, June 2006.

[4] M. Abusubaih, J. Gross, S. Wiethoelter, and A. Wolisz, "On Access Point Selection in IEEE 802.11Wireless Local Area Networks," in *WLN 2006*, November 2006.

[5] S. Vasudevan, K. Papagiannaki, C. Diot, J. Kurose, and D. Towsley, "Facilitating Access Point Selection in IEEE 802.11 Wireless Networks," in *Internet Measurement Conference 2005*, October 2005.

[6] S. Kandula, K. C.-J. Lin, T. Badirkhanli, and D. Katabi, "FatVAP: Aggregating AP Backhaul Capacity to Maximize Throughput," in *5th USENIX Symposium on Networked Systems Design and Implementation*, April 2008.

[7] D. Giustiniano, E. Goma, A. L. Toledo, I. Dangerfield, J. Morillo, and P. Rodriguez, "Fair WLAN Backhaul Aggregation," in *MobiCom10*, September 2010.

[8] H. Soroush, P. Gilbert, N. Banerjee, B. Levine, M. Corner, and L. Cox, "Spider: Improving Mobile Networking with Concurrent Wi-Fi Connections," in *SIGCOMM11*, August 2011.

[9] R. Chandra, P. Bahl, and P. Bahl, "MultiNet: Connecting to Multiple IEEE 802.11 Networks Using a Single Wireless Card," in *INFOCOM, 2004*, March 2004.

[10] J. Jun, P. Peddabachagari, and M. Sichitiu, "Theoretical maximum throughput of IEEE 802.11 and its applications," in *Network Computing and Applications'03*, April 2003.

[11] M. Bredel and M. Fidler, "Understanding Fairness and its Impact on Quality of Service in IEEE 802.11," in *INFOCOM 2009*, April 2009.

[12] J. Ha and C.-H. Choi, "TCP Fairness for Uplink and Downlink Flows in WLANs," in *Globecom'06*, November 2006.

[13] Yi Wu and Zhisheng Niu and Junli Zheng, "Upstream/downstream unfairness issue of TCP over wireless LANs with per-flow queueing," *WIRELESS COMMUNICATIONS AND MOBILE COMPUTING*, vol. VOL.5, May 2005.

[14] Eun-Chan Park and Dong-Young Kim and Hwangnam Kim and Chong-Ho Choi, "A Cross-Layer Approach for Per-Station Fairness in TCP over WLANs," *IEEE TRANSACTIONS ON MOBILE COMPUTING*, vol. VOL.7, NO.7, JULY 2008.

[15] Nicola Blefari-Melazzi and Andrea Detti and Ibrahim Habib and Alessandro Ordine and Stefano Salsano, "TCP Fairness Issues in IEEE 802.11 Networks: Problem Analysis and Solutions Based on Rate Control," *IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS*, vol. VOL.6, NO.4, APRIL 2007.