International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:4, No:1, 2010

Heterogeneity-Aware Load Balancing for Multimedia Access over Wireless LAN Hotspots

Yen-Cheng Chen and Gong-Da Fang

Abstract**—**Wireless LAN (WLAN) access in public hotspot areas becomes popular in the recent years. Since more and more multimedia information is available in the Internet, there is an increasing demand for accessing multimedia information through WLAN hotspots. Currently, the bandwidth offered by an IEEE 802.11 WLAN cannot afford many simultaneous real-time video accesses. A possible way to increase the offered bandwidth in a hotspot is the use of multiple access points (APs). However, a mobile station is usually connected to the WLAN AP with the strongest received signal strength indicator (RSSI). The total consumed bandwidth cannot be fairly allocated among those APs. In this paper, we will propose an effective load-balancing scheme via the support of the IAPP and SNMP in APs. The proposed scheme is an open solution and doesn't need any changes in both wireless stations and APs. This makes load balancing possible in WLAN hotspots, where a variety of heterogeneous mobile devices are employed.

Keywords**—**Wireless LAN, Load balancing, IAPP, SNMP.

I. Introduction

THE IEEE 802.11 [1] wireless local area network (WLAN) technology is known as one of the most prevailing wireless communication options in the recent years. Nowadays, many public hotspot areas, such as airports, coffee shops, and conference centers, have provided WLAN access to allow users to get connected to the Internet. Indeed, it is very popular and convenient for many people to use WLANs in their daily work and life. As public WLAN access becomes popular, it is very possible in a hotspot area that a lot of users access the Internet through the same WLAN simultaneously. A WLAN provides wireless access through an access point (AP). The bandwidth offered by an AP is limited and shared by all the mobile stations associated with the AP [2]. An IEEE 802.11g WLAN provides a transmission speed up to 54 Mbps (with typical data rate of 25 Mbps), which is shared by all the current users associated with the same AP. On the other hand, due to the more offered bandwidth in the Internet, more multimedia services emerge and much multimedia information can be accessed via the Internet. Indeed, there is a significant growth of real-time multimedia applications, such as video on demand (VOD) and voice over IP (VoIP). Therefore, it becomes popular that people access multimedia services through WLAN hotspots.

Yen-Cheng Chen is with the National Chi Nan University, Puli, Nantou, Taiwan (corresponding author to provide phone: +886-49-2910960 ext. 4654; fax: +886-49-2915205; e-mail: ycchen@ncnu.edu.tw).
Gong-Da Fang is with the Telecommunication Laboratories, Chunghwa Telecom. Co. Ltd., Taiwan (e-mail: gongda.fan@gmail.com).

Due to the limited bandwidth offered in WLAN hotspots, it is a challenge to provide multimedia access with acceptable quality of service especially when there are many current users in a hotspot. Indeed, many WLAN users have experienced poor multimedia quality or unstable wireless connections. Thus, wireless ISPs are often called to deal with the trouble of network congestion at certain popular place within the network. It is anticipated that Internet telephony and video communication applications through WLANs will be increasingly popular. Such congestion problem due to the scarce bandwidth in a WLAN will occur more frequently.

A possible solution to the WLAN congestion problem is the installation of multiple APs in the same WLAN hotspot. For example, installing two 802.11g APs in a WLAN hotspot can offer a bandwidth up to 108 Mbps theoretically. However, a mobile station (STA) usually gets associated with an AP with the strongest RSSI (Received Signal Strength Indication). The strength of a received signal is mainly determined by the distance between a mobile station and the AP. Therefore, it is possible that mobile stations are unevenly associated with two APs. As a result, the offered bandwidth cannot be equally shared by all the mobile stations. The congestion problem will still exist in a WLAN hotspot even after multiple APs are deployed.

This paper aims to develop a practical load-balancing scheme to improve the bandwidth utilization of a WLAN with multiple APs. For a fair use of bandwidth among APs, a new incoming mobile station should be associated with the AP with a maximum available bandwidth. As a result, the balance of bandwidth usage among APs can be achieved. Since in a public WLAN hotspot there are a number of mobile users with a variety of mobile devices and communication software to access the WLAN, a practical load-balancing scheme should be developed without the need of installing particular hardware and software in mobile devices. That is, the support of heterogeneity awareness in mobile devices is crucial to the design of a load-balancing scheme for public WLAN hotspots. In practice, heterogeneity-awareness in mobile devices is a challenging issue for a load-balancing scheme, because mobile devices, without any control, may try to connect to any AP at their will before the load-balancing scheme takes effect. Therefore, it is inevitable that a mobile device is initially associated with an AP with a higher bandwidth load. An effective heterogeneity-aware load balancing scheme should be aware of the unbalanced case as soon as possible and is capable of switching the association to another AP with lower load. These capabilities should be achieved without any particular control in mobile devices. For this, we will propose a load-balancing scheme on WLAN APs with the support of the IEEE 802.11F Inter-Access Point Protocol (IAPP) [3]. It will be shown that, by introducing a pseudo AP as an association dispatcher, IAPP can be used to monitor the association and re-association services in collocated APs. To realize the actual bandwidth consumption in each AP, the proposed scheme also uses the Simple Network Management Protocol (SNMP) [4] to collect the traffic statistics of each AP. Furthermore, SNMP is used to perform appropriate address filtering to enforce proper associations between mobile devices and APs. Both IAPP and SNMP are standard protocols supported in off-the-shelf enterprise-level WLAN APs. Hence, in addition to the heterogeneity-aware approach, the proposed scheme is an open solution to WLAN load balancing. An implementation of the proposed scheme will also be presented in this paper.

The remainder of this paper is organized as follows. Section II introduces previous related work. In Section III, we will describe the proposed load-balancing scheme based on IAPP and SNMP. In Section IV, we will present a real implementation of the proposed scheme. Experiments on the implementation are given in Section V. Finally, conclusions and future work are given in Section VI.

II. Related Work

Network congestion has been an important issue in wired networks. In the recent years, due to the popularity of public WLANs and the rapid growth of wireless multimedia applications, network congestion becomes a critical issue in WLAN hotspots. When a number of users access the Internet and run multimedia applications in the same WLAN hotspot, the bandwidth demands for multiple multimedia applications may reach or even exceed the maximum bandwidth that a single AP can offer. Network congestion thus happens whenever the bandwidth consumption reaches the network capacity of the WLAN.

A. Balachandran et al. [5] analyze the reasons of network congestion and propose a solution to congestion relief in WLAN environments. They show that the hotspot congestion can be relieved through explicit channel switching and network directed roaming. The idea of these solutions is to decide whether the network can provide the requested service in the user's current cell, or the network will direct the user to neighbor cells when a mobile station tries to connect to the network. This research concludes that distributing mobile stations among APs can improve the network utilization and throughput.

In [6] and [7], the authors study the dynamic load balancing issue in WLANs. Rather than just selecting the AP with the strongest RSSI value, they proposed an enhanced association policy for AP selection, when a mobile station wants to connect to a WLAN. The new policy considers the mean RSSI, number of associations that are in APs, and the signal-to-interference ratio. Experiments show that the new AP selection policy can lead to better network performance and lower network load. The study in [6] also suggests that each AP should support the Inter-Access Point Protocol for the automatic selection of an operation channel in the initial state of an AP, so as to reduce the signal interference between APs and mobile stations. However, the proposed scheme needs to add new fields in Beacon frames and Probe Response frames to carry additional information needed by the proposed algorithm. This approach requires a modification of the current IEE 802.11 protocol.

In [8], Y. Bejerano et al. proposed an algorithmic solution to determine the user-AP associations that ensure a network-wide max-min fair bandwidth allocation. The solution is proposed to provide fair services to the users and to balance the load among APs. By analyzing the associations between mobile stations and APs, they provide a rigorous formulation of the association control problem that considers bandwidth constraints of both the wireless and backhaul links. Their formulation shows the strong correlation between the fairness and load balancing, and they apply load balancing techniques to obtain a near optimal max-min fair bandwidth allocation. The optimal max-min fair bandwidth allocation could maximize the overall system throughput.

In [9], H. Velayos et al. proposed a scheme that introduces a Load Balancing Agent (LBA) in each AP. Each LBA periodically broadcasts the load level of its AP to the common backbone. Using the reports from other LBAs, each LBA assesses whether the load is balanced among neighboring APs.

Yen et al. [10] propose an SNMP-based approach, in which standard MIB-II objects are used in the estimation of AP utilization. The utilization, together with the number associated stations, is further used to calculate the normalized residual bandwidth, as the load metric of the proposed scheme. A benefit of the proposed approach is its practical implementation using off-the-shelf WLAN devices. However, each wireless station should additionally install a client software for load balancing.

The above studies investigate WLAN load-balancing from various aspects. In summary, the key issue to load balancing for a multimedia service performed by an incoming STA is to select the lowest-load AP for a new association. As the AP selection process proceeds as STAs enter the WLAN hotspot, the bandwidth load can be nearly balanced among APs. Many researchers define different load metrics to select the lowest-load AP. As indicated in a survey [11], load metrics and load-balancing methods are the two primary issues in these studies. The survey gives a summary of possible load-balancing metrics: number of wireless stations[12], channel utilization [13, 14], frame drop rate[15], queuing and contention delay [16], throughput [9,12], and bandwidth usage [8, 10, 16]. Currently, to our best knowledge, it is still unknown what is the best load indicator to quantify the load in WLANs, since there are various factors affecting the quality of WLANs. Furthermore, some load metrics are hard to measure in practice, or not supported in off-the-shelf WLAN products. In addition to the load metric issues, it is another challenge to implement proposed WLAN load-balancing methods. The survey [11] also indicates that either wireless station or AP requires change to accommodate load-balancing. This brings difficulties for a real implementation of load-balancing schemes in public WLAN hotspots. In this paper, we will face the challenge and will develop a novel load balancing scheme, without any changes in both the wireless station and the AP, for use in WLAN hotspots.

In addition to the above theoretical researches, there have been WLAN products incorporating load-balancing features in their network device drivers, AP firmware and WLAN cards [3, 8]. However, these proprietary solutions usually require that APs and the wireless cards in mobile stations be of the same vendor. Therefore, these solutions cannot be applied in public WLAN hotspots, where there are a variety of mobile stations of different vendors.

III. The Proposed Scheme

Many concurrent multimedia accesses through the same WLAN hotspot may lead to network congestion due to insufficient bandwidth. The hotspot congestion can be relieved by installing multiple APs within the same area. A mobile station usually associates to the AP with the strongest RSSI. On the other hand, current APs accept all the association requests if the requesting mobile stations are authorized to access the WLAN. As a result, the associations may lead to an unbalanced bandwidth use among APs. The proposed scheme is developed based on an adaptive association control to enforce that each mobile station is only allowed to be associated with the AP with less bandwidth consumption. Accordingly, that bandwidth consumption can be evenly distributed among APs. For the adapted association control, the following requirements are necessary for the proposed scheme:

(1).**Association Monitoring**: The first prerequisite for the adaptive association control is that the load balancing scheme should be aware of the occurrence of all associations in APs. Thus, the scheme may further perform an appropriate association control over the associations.

(2).**Association Control**: The load balancing scheme should be able to guide the APs to accept or reject an association such that associations can be evenly distributed to APs. If an association is rejected, the disassociated mobile station can then properly associate with the AP which less bandwidth use.

(3).**Low Handoff Overhead**: If a mobile station is rejected to associate to an AP due to the unbalance of bandwidth use, the mobile station will try to make an association with another AP. The overhead of the forced handoff should be kept minimal.

(4).**Traffic Monitoring**: To properly perform association control, the load balancing scheme should be aware of the actual bandwidth consumption so far in each AP.

The proposed scheme is developed mainly based on association control using IAPP. IAPP was developed to support the DS (Distributed System) functions defined in IEEE 802.11. Although IAPP doesn't address how to improve the overall bandwidth utilization in a DS, we find that several characteristics of IAPP meet the above first three requirements for the load balancing scheme. We will implement an association dispatcher as a pseudo AP which talks with real APs via IAPP. The use of IAPP in the proposed scheme will be described later. For effective traffic monitoring, we will use SNMP to gather traffic statistics data for use in determining the AP with the lowest load. The *ifInOctets* and *ifOutOctets* objects of *ifTable* in MIB II [17] are retrieved periodically to provide such information. SNMP will also be used in association control. Through MAC address filtering via SNMP, a new incoming mobile station is only allowed to access the APs with lowest load. The *dot1dStaticTable* objects in Bridge MIB [18] are used for MAC address filtering.

Figure 1 shows the system architecture of the proposed scheme. Without loss of generality, we assume that there are two APs collocated in a WLAN hotspot and both APs are linked to a Fast Ethernet switch, with a bandwidth of at least 100 Mbps in each port. In the wired LAN which the switch is connected to, an Association Dispatcher Server (ADS) is deployed to implement the association dispatcher function. In addition, a media streaming server is located in the LAN to provide a streaming multimedia service. The media streaming server can also be placed in any network where the WLAN can reach. If the media streaming server is located elsewhere, it is assumed that the path from the media streaming server to the APs can offer sufficient bandwidth for streaming media access.

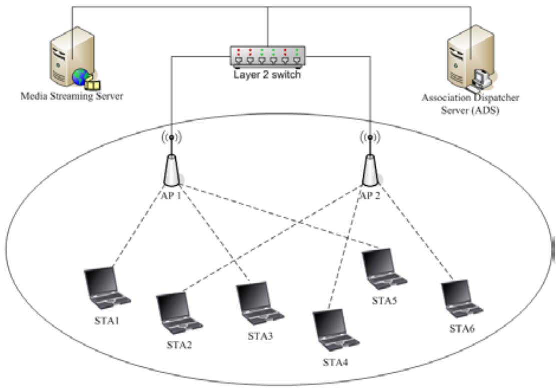


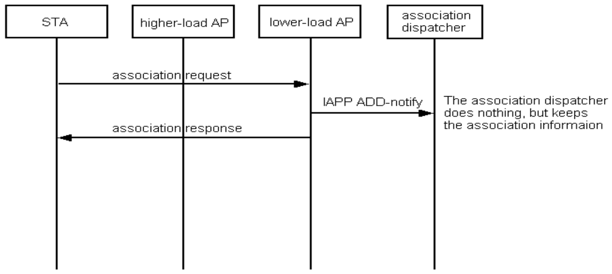
Fig. 1 System architecture of the proposed scheme

In the scheme, both APs as well as ADS are IAPP-enabled. ADS can be regarded as a pseudo AP which is involved with the two real APs to corporately support the DS function. In fact, both real APs behave as usual, as if there is another AP, the pseudo AP, providing the same wireless access service in the DS. ADS will use IAPP to monitor all associations and to perform association control over the associations in APs. In addition, ADS also acts as an SNMP manager of the two APs. ADS will periodically use the SNMP Get requests to retrieve the traffic statistics data of the wireless interface in each AP.

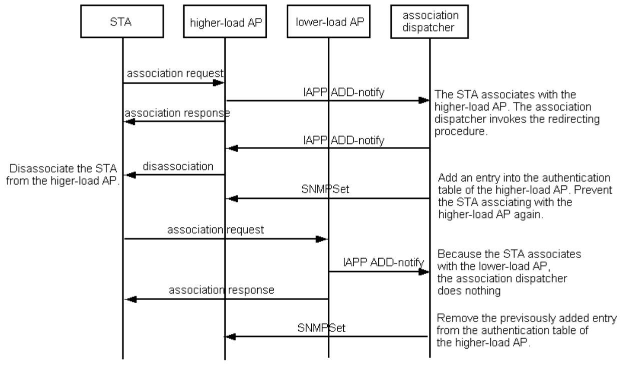
The ADS, as a pseudo AP, always listens to the wired LAN to receive IAPP packets. Before the association dispatcher can take control of the associations, a new arriving STA would first become associated with one of the real APs. Since each AP will multicast an IAPP ADD-notify packet upon receiving an association request from a STA, ADS will receive the IAPP

ADD-notify packet. This packet triggers the dispatch decision procedure of the association dispatcher in ADS. By examining each received ADD-notify packet, the association dispatcher will know which AP a STA is now associated with. If the STA is associated with an AP with a lower-load in bandwidth use, the association dispatcher in ADS will do nothing but keep the association information, as shown in Figure 2.(a). Otherwise, the association dispatcher will ask the AP to abort the association immediately such that the STA can establish another association with the other AP. To achieve this, the association dispatcher will emit another ADD-notify packet which contains the MAC Address of the STA. Sending this packet is to announce that a new association has been established between the STA and the pseudo AP, i.e. the ADS. Recall that IAPP enforces that only one association is allowed for each STA at the same time. Each AP receiving the ADD-notify packet will check if the MAC address indicated in the IAPP ADD-notify packet exists in the association table of the AP. If exists, the AP will remove the corresponding association entry and disassociate with the STA. This action breaks the association between the STA and the AP with a higher bandwidth use. As a result, the association dispatcher can take advantage of IAPP to effectively prohibit association requests in the AP with a higher bandwidth use. In general, when a STA fails to make an association with an AP, the STA will look for another AP to establish a new association. Therefore, by our scheme, the STA will find the lower-load AP and successfully make an association with it. In practice, due to the stateless property of some WLAN products, it is possible that a STA repeatedly tries to request an association with the same AP, i.e. the higher-load AP, even if the STA was disassociated from the AP previously. In the case, the above scenario of association dispatches will repeat persistently. To avoid the repeated situation, we enable the MAC Address authentication scheme in the AP to temporarily disallow any association request from the STA. The MAC Address authentication scheme can be configured via setting the *dot1dStaticTable* objects in Bridge MIB. Figure 2.(b) shows the sequences of the proposed scheme when a STA associates with a higher-load AP.

In fact, the proposed scheme allows the use of other metrics to evaluate the network load, such as the mean RSSI, the number of STAs that currently associate with an AP, the CPU usage of an AP, and so on. Many researchers also proposed improved criteria to assess the load of the network, but some criteria are hard to measure or are too complicated to implement. Our study focuses on the balanced use of the bandwidth among APs to probably prevent congestion due to the unfair use of bandwidth. Therefore, we choose the bandwidth utilization of the WLAN as the criterion to select APs. An advantage of our approach is the ease of measuring bandwidth usage. Almost all enterprise-level APs support SNMP for remote management. The *ifInOctets* and *ifOutOctets* objects of *ifTable* in MIB II provide the inbound and outbound traffic statistics of APs for the estimation of bandwidth utilization. In ADS, there is a remote traffic monitoring program responsible for gathering



(a). Sequence diagram: STA associates with a lower-load AP



(b). Sequence diagram: STA associates with a higher-load AP

Fig 2. Sequence diagrams of the proposed scheme

traffic statistics data from APs. The traffic monitoring program is run periodically with polling interval T.

Assume the maximum wireless LAN bandwidth of the i-th AP is B_{imax} and the AP can offer at most S_{imax} STAs to play high quality multimedia streaming video. Therefore, we define IU_i (Increment Unit) as the pre-reserved bandwidth for an incoming STA:

$$IU_i = \frac{B_{i\max}}{S_{i\max}}$$

IU_i denotes an estimated bandwidth consumed by a STA for the transmission of multimedia data with acceptable QoS. In our scheme, IU_i also denotes a predicted traffic increment after a newly arrived STA associates with the i-th AP. Since the polling interval of traffic monitoring is T, the traffic statistics data for the evaluation of available bandwidth will be updated every T seconds. Suppose $B_i(t)$ is the bandwidth consumption of the i-th AP measured at the t-th polling, and $N_i(t)$ is the number of newly arrived STAs associated with the i-th AP so far after the t-th polling, The available bandwidth of the i-th AP within the T seconds after the t-th polling, denoted by $A_i(t)$, is estimated as follows.

$$A_i(t) = B_{imax} - B_i(t) - N_i(t)*IU_i$$

After an incoming STA first gets associated with the i-th AP and an IAPP Add-notify packet is sent to the association dispatcher, the association dispatcher will determine whether $A_i(t)$ is maximum among APs. If yes, the association dispatcher does nothing but increases $N_i(t)$. Otherwise, an association control by IAPP follows. Accordingly, the STA will be only allowed to associate with AP with maximum estimated available bandwidth.

IV. Implementation

Currently, there are a few off-the-shelf AP products which support the IEEE 802.11F IAPP protocol. The APs used in our implementation are SMC 2555W-AG APs, which support the draft version of the IAPP standard. The difference between the draft version and the standard version is the method of transmitting IAPP ADD-notify packets. In the draft version, IAPP ADD-notify packets are sent via broadcast, while multicast is used in the standard one. Therefore, in our implementation the ADS must be located in the same subnet of APs to receive IAPP ADD-notify packets. In the implementation, we put two identical APs within the same location area, and install the ADS in the wired LAN which the two APs are connected to.

We use AdventNet JAVA SNMP API [19] to implement our scheme. All our implementation is in the ADS. Figure 3 shows the program architecture of the ADS.

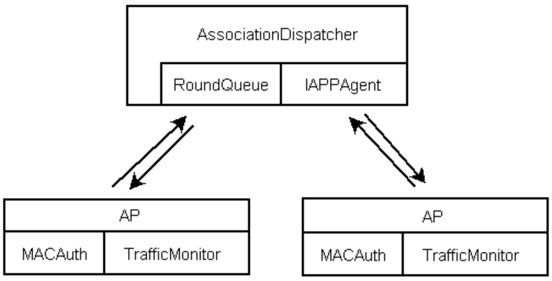


Fig. 3 Program architecture of ADS

There are three primary classes in the ADS: *AssociationDispatcher*, *AP*, and *IAPPAgent*. The *AssociationDispatcher* is the main component of this system. The *AssociationDispatcher* class is responsible for controlling the procedure of the program, collecting the traffic statistics from the APs periodically, evaluating the available bandwidth of each AP, and determining whether a STA needs to be switched to the lower-load AP. The traffic statistics are collected periodically from the *AP* objects in the system. Each *AP* object, corresponding to a real AP, uses an SNMP Get request to retrieve traffic statistics data from the corresponding AP. The *AssociationDispatcher* object interacts with the *AP* objects to get traffic statistics data for further processing. Two objects are included in the *AP* class. The *TrafficMonitor* object is used to collect the traffic statistics from real APs, and the *MACAuth* object is used to configure the MAC address authentication table of each AP. The *IAPPAgent* is responsible for receiving and generating ADD-notify packets. The *IAPPAgent* extracts the MAC address and sequence number in the payload of the received ADD-notify packet and uses these values to generate another ADD-notify packet. The *RoundQueue* is a queue structure designed to buffer the traffics data of each AP.

V. Experiments

A. Experiment Settings

The experiment environment is set up as Figure 1. There are two SMC2555W-AG APs, six notebook PCs acting as STAs, a streaming server, and the Association Dispatcher Server (ADS). The streaming server is a Pentium 4 PC server running Windows 2003 server and Windows Media Server software. The ADS is a Pentium 4 PC server running the association dispatcher program. Table I lists the parameters used in the experiments.

TABLE I
System Parameters of Experiments

Parameter	Value
B_{imax}	11 Mbps
S_{imax}	20
Bit Rate of Streaming Video	384 Kbps, 600Kbps
Polling interval for traffic measurement	15 seconds, 1 minute
Transmission Protocol	MMS(TCP), MMS(UDP)
Streaming Client	Windows Media Player 9 Windows Media Player 10

The STAs are running Windows XP. The user of each STA is able to specify a preference list of APs. Each STA tries to associate with APs in the order of the APs specified in the preference list. The two APs in our experiments are added into the preference list. We use the Windows Media Encoder to generate 384 Kbps and 600 Kbps sample videos. These videos are stored in the streaming server. The Fast Cache feature is disabled. Each streaming video is associated with an Advanced Stream Redirector (ASX) file used to redirect Windows Media Player to connect to the correct Windows Media Server. The ASX files are put in a web server, and any user can access the streaming service by opening an ASX file [20]

The Windows Media Player in each STA communicates with the streaming server via the Microsoft Media Server (MMS) protocol. The MMS has a feature called *protocol rollover* [20]. When a Windows Media Player connects to the streaming server via the MMS protocol, the protocol rollover will choose the optimal transmission protocol according to the Media Player version and current network environment. The protocol rollover chooses protocols in the following order:

1. RTSPU (RTSP using UDP)
2. RTSPT (RTSP using TCP)
3. HTTP

These protocols transmit multimedia data through unicast. Most personal multimedia streaming services in the Internet are transmitted by unicast protocols. Therefore, we choose the unicast MMS protocol to provide the streaming services.

B. Balancing Index of Bandwidth Load

There are many aspects of measuring the performance of our load balancing scheme. Since we concern the fairness of the bandwidth allocation, the concept of the balance index, introduced in [5, 9, 21], is adopted to evaluate the fairness of bandwidth use among APs. The balance index β in terms of bandwidth usage $B_i(t)$ is defined as follows:

$$\beta = \frac{(\sum B_i(t))^2}{n(\sum B_i(t)^2)}, for\ i = 1, 2, \ldots, n$$

In the context of our scheme, n denotes the number of APs, and $B_i(t)$ is the bandwidth consumed in the i-th AP at the t-th polling. The balance index is computed according to the real traffic measurement via SNMP. The balance index has the following properties [21]:

a. The value of β is bound between 0 and 1 (0% and 100%). β reaches 1 when a fair bandwidth allocation is achieved. On the other hand, a smaller β indicates a more unfair bandwidth allocation.
b. The balance index is independent of scale. That is, the balance index can be calculated regardless of the measurement unit.
c. The balance index is a continuous function. It can reflect any slight change in bandwidth allocations.
d. If we have n APs but the bandwidth is contributed by only k APs, this unfair bandwidth allocation will lead to a balance index with value near k/n.

C. Experimental Scenarios

We design three scenarios to simulate user behaviors and network traffic characteristics in a hotspot environment. We assume that all STAs first choose AP 1 to associate with to result in an unfair condition in associations. Then, we observe the effectiveness of our scheme in the three scenarios.

Scenario 1: Gradually Join
At the beginning of every minute, a new STA arrives and attempts to associate with AP 1. After the association is established, the STA starts to play streaming multimedia from the streaming server. The traffic of APs is measured continuously for 5 minutes after all STAs get associated.

Scenario 2: Join Together
All of the STAs associate with AP 1 within the first minute. When the STAs complete the associations, they will start to play streaming videos. We will observe the traffic of APs for 5 minutes. This scenario is to test the effect of the bandwidth load estimation by periodical polling of real traffic.

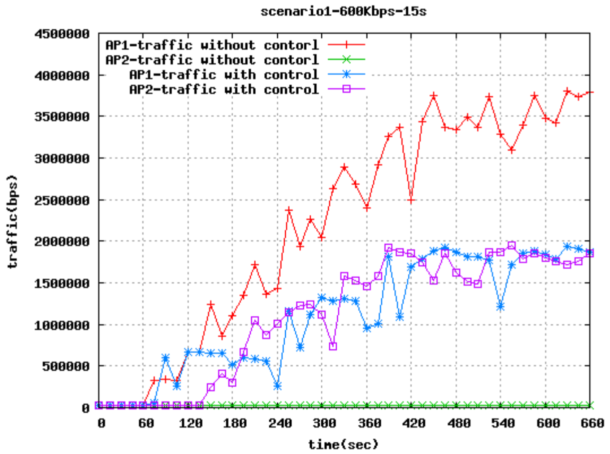
Scenario 3: Gradually–then-together Join
Two of STAs associate with AP 1 at the beginning of every minute. The other four STAs will associate with AP 1 at 3rd minute. All of the STAs would start to play streaming videos after the STAs complete the associations. We also observe the change of traffic of APs for 5 minutes.

D. Experimental Results

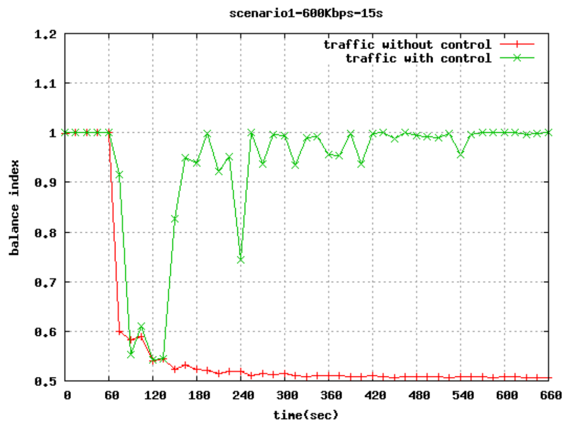
In each scenario of the experiment, we measured the consumed bandwidth in each AP and evaluated the balance index for 384 Kbps and 600 Kbps streaming videos. There are two polling intervals of traffic statistics data: 15 seconds and 60 seconds. The first minute is used to observe the background traffic of the network when there are no STAs associated with APs. The experiment will start at 2nd minute. In the following, only the experimental results of 600 Kbps streaming videos with 15 sec polling interval are presented. The other experiments have similar results.

Scenario 1:

The experimental result of scenario 1 is shown in Figure 4. Since STAs associate with AP 1 one station per minute in scenario 1, all the figures show that without any association control the monitored traffic of AP 1 increase gradually but the traffic of AP 2 remains unchanged. These figures show that with association control the traffic is almost balanced between APs, because half of STAs are forced to associate with AP 2. All the balance index values also show the superiority of our load balancing scheme. Our experiment shows that by association control the balance index increases from 52% to 99%. In Figure 4.(b), the balance index suddenly falls to 74% at the 240-th second, because there are three STAs, one in AP 1 and two in AP 2, resulting in a temporarily unfair use of bandwidth between APs.



(a). Traffic

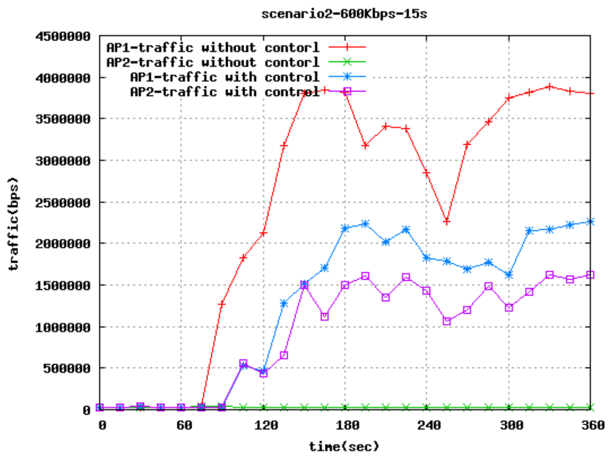


(b). Balance index

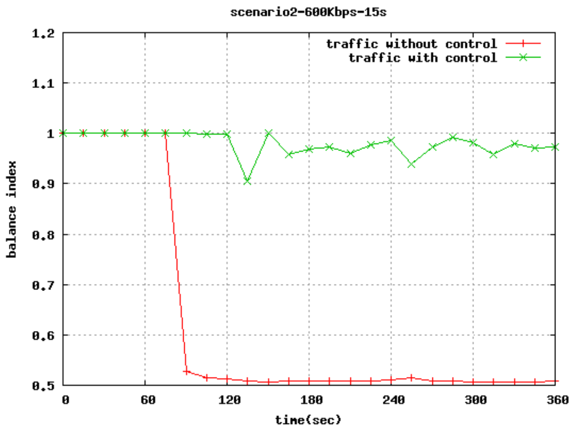
Fig. 4 Experimental result of scenario1

Scenario 2:

In the experiment of scenario 2, all of the STAs associate with AP 1 within the second minute. The experimental result of scenario 2 is shown in Figure 5. In the absence of association control, all STAs associate with AP 1. As a result, as shown in this figure, the traffic in AP 1 increases significantly but the traffic in AP 2 remains unchanged. With the association control, all STAs are first associated with AP 1 within the second minute, however immediately half of STAs are forced to associate with AP 2 before they play multimedia videos. Therefore, traffic is almost balanced between APs. This experiment shows that our scheme can proactively control associations before unbalance occurs. As shown in Figure 5.(b), the balance index values of the proposed scheme are near 1 most of time, and the proposed scheme improves the balance index from 51% to 97%. In Figure 5.(a), the traffic curve of AP 1 without association control suddenly falls at the 255-th second, because network congestion occurs.



(a). Traffic

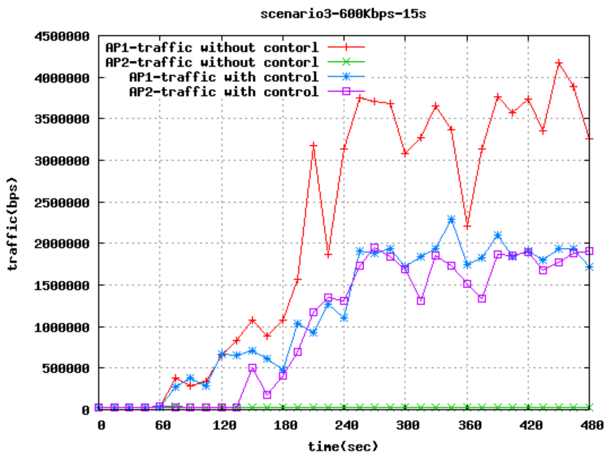


(b). Balance index

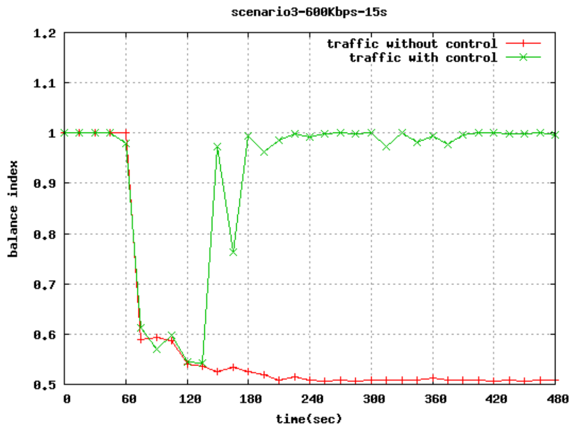
Fig. 5 Experimental result of scenario2

Scenario 3:

In the experiment of scenario 3, two of STAs associate with AP 1 at the first and second minute respectively, and the other four STAs associate with AP 1 within the third minute. The experimental result of scenario 2 is shown in Figure 6. In the absence of association control, the traffic of AP 1 increases gradually within the first two minutes and increases greatly after the third minute, resulting in the decrease of balance index values to near 50%. By our load balancing scheme, the STA arriving at the second minute is forced to associate with AP 2, resulting in the increase of balance index. Then, by association control, the other four STAs, arriving during the third minute, evenly associate with the two APs, resulting in the fair increase of traffic in both APs. As a result, the proposed load balancing scheme achieves higher balance index. Indeed, the balance index is improved to 99%. In this experiment, the balance index at the 120-th second doesn't show the effect of the proposed scheme because only one STA associates with AP 1 and therefore only the traffic of AP 1 increases, resulting in a balance index with value 54%, the same as the case without association control. In Figure 6.(a), the traffic curve of AP 1 without association control suddenly falls at the 360-th second because the network is in congestion.



(a) Traffic



(b) Balance index

Fig. 6 Experimental result of scenario3

E. Discussion and Analysis

In the above experiments, we also measure the handoff latency of STAs when they are being redirected from the higher-load AP to the lower-load one. A wireless packet sniffer is deployed to observe the handoffs. The wireless packet sniffer is a notebook running Linux and equipped with a D-link DWL-G650 PCMCIA card. By referring [22, 23], we use Wireshark [24] to capture the layer 2 wireless frames and measure the handoff latency of the STAs.

The handoff latency is measured by the time difference between receiving an IEEE deauthentication frame to receiving an IAPP ADD-notify packet. The wireless association between a

STA and an AP would be terminated when the STA receives an IEEE 802.11 deauthentication frame from the AP. To continue the wireless connection, the STA may try to associate with the same AP for a number of times until the STA gives up the trials. Then, the STA tries to find another AP to associate with. If the STA successfully establishes a new association with another AP, an IAPP ADD-notify packet will be sent by the AP. The IAPP ADD-notify packet is also propagated through the IEEE 802.11 WLAN and will be captured by the wireless packet sniffer. Therefore, we can measure the handoff time in the wireless packet sniffer. We find that the handoff times are different among WLAN network interface cards (NICs) of different vendors. The measured results for four different NICs are summarized in the Table II.

Most handoff time is contributed by an STA when it tries to associate with the same AP. The difference in handoff times is caused by different parameter values adopted in wireless card drivers. Each wireless card was designed with different performance considerations.

TABLE II
HANDOFF LATENCY OF WIRELESS CARDS

Wireless NICs	Handoff latency (sec)
D-Link Air DWL-650	5.4557
Spark WL-311F	7.0119
D-Link DWL-G122	13.8475
D-Link DWL-G650	15.2615

IEEE 802.11F IAPP standard also provides the CACHE-Notify service for APs to transfer context information blocks to neighboring APs. The context information might contain the authentication information of STAs to achieve fast handoff. The IAPP standard also provides a proactive caching algorithm to build a neighbor graph. APs can effectively transfer context information according to the neighbor graph. In addition, A. Mishra et al. [25] also study the neighbor graph algorithm and show simulation results that the handoff latency can be reduced. Furthermore, in [26, 27, 28], researchers develop the packet buffering mechanism and the zero-delay signaling mechanism to help fast handoff in WLANs These schemes are achieved by using IEEE 802.11F IAPP.

In the proposed scheme, redirecting a STA from the higher-load AP to the lower-load one occurs before the STA receives a streaming video. Although the STAs would spend a longer handoff time to associate with the lower-load AP, once the association is established the user of the STA is able to get more bandwidth to enjoy streaming videos.

VI. CONCLUSIONS

We have proposed a load balancing scheme using IEEE 802.11F IAPP to balance the bandwidth consumption in WLAN hotspots where multiple APs are deployed. This scheme can effectively relieve the network congestion and accommodate more users to enjoy streaming videos simultaneously in a hotspot environment. Different from other theoretical researches, a major merit of our scheme is that the scheme can be realized using current WLAN products. This study has also shown that the proposed scheme is easy to deploy. Only APs with the support of IAPP are required, in addition to the Association Dispatcher Server installed within the same subnet of the APs. On the other hand, no particular hardware or software is required in STAs. Therefore, the proposed scheme is suitable for a hotspot environment, where multi-vendor STAs access the WLAN without any change of the configuration in STAs.

The performance of the proposed scheme has been evaluated by real experiments, instead of simulations. The experimental results have shown that the bandwidth consumption can be fairly distributed among APs and the balance index in the WLAN can be significantly improved from 51% to 99%. Therefore, the proposed scheme is an effective solution to load balancing.

In the future work, we will investigate the effective solution to reduce the handoff latency and develop a better way to assess the load of WLAN. In addition, we will enhance the proposed scheme with admission control. That is, when the bandwidth cannot offer more STAs to play multimedia applications with acceptable QoS, any newly arriving STA will not be allowed to associate with any AP until the bandwidth is released by other leaving STAs.

REFERENCES

[1] IEEE 802.11 WG, “IEEE 802.11 Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications,” IEEE Standard, 1999.
[2] Y. Moritani, Y. Atsumi, “Efficient Bandwidth Utilization through Multicast Receiver Aggregation in wireless LAN spot,” The 11th IEEE International Conference on ICON2003, pp. 579-584, 28 September -1 October 2003.
[3] IEEE 802.11 WG, “802.11F IEEE Trial-Use Recommended Practice for Multi-Vendor Access Point Interoperability via an Inter-Access Point Protocol Across Distribution Systems Supporting IEEE 802.11 Operation,” IEEE Standard, 14 July 2003.
[4] J. D. Case, M. Fedor, M. Schoffstall, C. Davin, “Simple Network Management Protocol (SNMP),” RFC 1098. April 1989.
[5] A. Balachandran, P. Bahl, and G. M. Voelker, “Hot-Spot Congestion Relief in Public-area Wireless Networks,” In Proc. Fourth IEEE Workshop on Mobile Computing Systems and Applications, pp. 70-80, June 2002.
[6] I. Papanikos and M. Logothetis, “A study on dynamic load balance for IEEE 802.11b wireless LAN,” In Proc. COMCON, 2001.
[7] L. Nuaymi, J. El-Sayah, “Access Point Association in IEEE 802.11 WLAN,” In Proc. Information and Communication Technologies: From Theory to Applications (ICTTA), April 2004.
[8] Y. Bejerano , S. J. Han , L. (Erran) Li, “Fairness and load balancing in wireless LANs using association control,” In Proc. the 10th annual international conference on Mobile computing and networking, September 26-October 01, 2004, Philadelphia, PA, USA.
[9] H. Velayos, V. Aleo and G. Karlsson, “Load balancing in overlapping wireless LAN cells,” In Proc. IEEE ICC 2004, Paris, France, June 2004.
[10] L.-H. Yen and T.-T. Yeh, “SNMP-Based Approach to Load Distribution in IEEE 802.11 Networks,” Proc. Vehicular Tech. Conf., vol. 3 (VTC2006-Spring), IEEE Press, 2006, pp. 1196–1200.
[11] Li-Hsing Yen, Tse-Tsung Yeh, and Kuang-Hui Chi, "Load balancing in IEEE 802.11 networks," IEEE Internet Computing, Vol. 13, pp. 30-38, Jan-Feb, 2009.
[12] Y. Fukuda and Y. Oie, “Decentralized 1. Access Point Selection Architecture for Wireless LANs: Deployability and Robustness,” Proc. Vehicular Tech. Conf., vol. 2 (VTC2004-Fall), IEEE Press, 2004, pp. 1103–1107.
[13] IEEE 802.11 WG, "Draft Supplement to Part 11: Wireless Medium Access Control (MAC) and physical layer (PHY) specifications: Medium Access Control (MAC) Enhancements for Quality of Service (QoS)", IEEE 802.11e/D13.0, Jan. 2005.

[14] M.T. Lee, L.-T. Lai, and D. Lai, "Enhanced Algorithm for Initial AP Selection and Roaming," US patent 0039817, Patent and Trademark Office, Feb. 2004.
[15] O. Brickley, S. Rea, and D. Pesch, "Load Balancing for QoS Enhancement in IEEE 802.11e WLANs Using Cell Breathing Techniques," Proc. IFIP Mobile and Wireless Communication Networks Conf., Int'l Federation for Information Processing, 2005.
[16] S. Vasudevan et al., "Facilitating Access Point Selection in IEEE 802.11 Wireless Networks," Proc. Internet Measurement Conf., Usenix Assoc., 2005, pp. 293–298.
[17] K. McCloghrie and M. Rose, "Management Information Base for Network Management of TCP/IP-based internets: MIB-II," RFC1213, March 1991.
[18] E. Decker, P. Langille, A. Rijsinghani, K. McCloghrie, "Definitions of Managed Objects for Bridges," RFC 1493, July 1993.
[19] AdventNet Java SNMP API, http://snmp.adventnet.com/
[20] Windows Media Services Help Document, Microsoft Corporation, 2005.
[21] D. M. Chiu and R. Jain, "Analysis of the Increase and Decrease Algorithms for Congestion Avoidance in Computer Networks," Journal of Computer Networks and ISDN, vol. 17, June 1989. [15] V. Aleo, "Load distribution in IEEE 802.11 cells", Master of Science Thesis, KTH, Royal Institute of Technology, March 2003.
[22] J. Yeo, M. Youssef, and A. Agrawala, "A Framework for Wireless LAN Monitoring and Its Applications," In Proc. the 2004 ACM workshop on Wireless security, October 2004.
[23] A. Mishra, M. Shin, and W. Arbaugh, "An Empirical Analysis of the IEEE 802.11 MAC Layer Handoff Process," ACM SIGCOMM Computer Communication Review, vol.33, no. 2, pp. 93-120, April 2003.
[24] Wireshark, http://www.wireshark.org/.
[25] A. Mishra, M. Shin, and W. Arbaugh, "Context Caching Using Neighbor Graphs for Fast Handoffs in a Wireless Network," In Proc. IEEE INFOCOM, March 2004.
[26] C. T. Chou and K. G. Shin, "An Enhanced Inter-Access Point Protocol for Uniform Intra and Intersubnet Handoffs," IEEE Transactions on Mobile Computing, vol. 04, no. 4, pp. 321-334, July/August, 2005.
[27] I. Samprakou, C. J. Bouras, and T. Karoubalis, "Fast and Efficient IP Handover in IEEE 802.11 Wireless LANs," 2004 International Conference on Wireless Networks (ICWN' 04), Las Vegas, Nevada, USE, June 2004.
[28] I. Samprakou, C. J. Bouras, and T. Karoubalis, "Improvements on "IP-IAPP": A fast IP Handoff Protocol for IEEE 802.11 wireless & mobile clients," WINET Journal, 2006.

Yen-Cheng Chen Yen-Cheng Chen received the Ph.D. degree in computer science from the National Tsing Hua University, Taiwan, in 1992. He was an associative researcher of the ChungHwa Telecom Labs. from 1992 to 1998. From 1998 to 2001, he was an assistant professor of the Department of Information Management, Ming Chuan University, Taiwan. Currently, he is an associate professor of the Department of Information Management, National Chi Nan University, Taiwan. His current research interests are network management, wireless networks, and security.

Gong-Da Fang Gong-Da Fang received the B.S. and M.S. degree in information management from National Chi Nan University, Taiwan, in 2004 and 2006, respectively. Currently, he is an assistant researcher of the Telecommunication Laboratories, Chunghwa Telecom. Co. Ltd., Taiwan. His interests include network management and Internet technologies.