

Cooperative content dissemination in multi-channel WLAN hotspots

Kate Ching-Ju Lin, Cheng-Fu Chou *

National Taiwan University, Institute of Networking and Multimedia, No. 1, Sec. 4, Roosevelt Road, Taipei 10617, Taiwan, ROC

Received 21 January 2007; received in revised form 7 November 2007; accepted 11 November 2007

Available online 21 November 2007

Responsible Editor: Prof. A. Capone

Abstract

It is common for hotspots in airports, cafes, and malls to deploy info-stations on their access points to provide local information such as flight schedule, multimedia access, sales and discount deals. The dissemination service, however, is limited by the wireless throughput in hotspots and cannot scale to a large number of users. This paper designs a new architecture that significantly improves the throughput of info-stations. We observe that usually the number of collocated info-stations is smaller than the number of orthogonal channels in IEEE 802.11, leaving some orthogonal channels unused, and thus wasting their capacity. We also observe that many users of an info-station download the same objects. Thus, we advocate an architecture, called hybrid-WLAN (H-WLAN), where mobile nodes can retrieve objects (multimedia content or information) from the info-station in the infrastructure mode, or exploit the idle 802.11 channels to cooperatively share their content with neighboring nodes in an ad-hoc manner. Our H-WLAN contains two components. The first component directs each object request to a nearby ad-hoc node that has the requested object, while avoids overloading any of the nodes. The second component balances the load across channels taking into account the popularity of the disseminated objects. We evaluate our design via simulations. The results show that our H-WLAN improves channel utilization and provides significantly higher throughput.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Info-station systems; Wireless hotspots; Cooperative data dissemination

1. Introduction

Wireless local area networks have become increasingly widespread in the last decade. Advances in wireless technology and the demand for mobile

Internet access have encouraged the development of hotspot technology [1,2]. Hotspots are wireless access points deployed in some public locations, such as airports, restaurants, travel agents, conference centres, cafes, and train stations. They can even be deployed in mobile environments [3,4], such as trains, subways, and buses. Content dissemination is one of the most common applications in hotspots. Recently, many info-station systems [5–7] have been

* Corresponding author. Tel.: +886 2 33664888x517.

E-mail addresses: cjlin@cmlab.csie.ntu.edu.tw (Kate C.-J. Lin), ccf@csie.ntu.edu.tw (C.-F. Chou).

set up in hotspots to disseminate objects (popular information or content) for mobile users. An object distributed by info-stations is usually *geographically-sensitive* (e.g., course slides in a class, travel information at metro-stations, and on-line documents in a conference) and *temporally-sensitive* (e.g., movie schedules and restaurant coupons). Since users in hotspots have some common interests in geographically- and temporally-sensitive objects, the popularity of the objects in an info-station is typically skewed. To take into account the above properties of info-station systems, we propose an efficient scheme that facilitates cooperative content dissemination in WLAN hotspots.

Conventional info-station systems usually operate as APs, and act as servers to provide a conduit to the Internet for mobile users. Each mobile node associates itself with one of APs within its transmission range to retrieve objects or access the Internet. Standard IEEE 802.11 supports multiple orthogonal channels; for example, there are three non-overlapping channels in 802.11b/g and eight non-overlapping channels in 802.11a. Hence, in a hotspot, APs in the vicinity of each other are usually allocated different non-overlapping channels such that they can operate simultaneously without interference. However, without loss of generality, a commercial AP is usually equipped with one transceiver, and can only operate over a single channel. Even though all APs in the coverage range of a hotspot are allocated non-overlapping channels, there are still some channels unused when the number of APs is smaller than the number of available channels. Thus, based on this observation, we develop a framework, called Hybrid-WLAN (H-WLAN), that can utilize bandwidth capacity of the idle channels, which are not allocated to APs, to improve the throughput of content dissemination in info-station systems.

Mobile nodes gathered around a hotspot usually have similar interests because the objects provided by an info-station are geographically- and temporally-sensitive. Therefore, mobile nodes can share their objects in a peer-to-peer manner without going through the info-station. In IEEE 802.11 wireless networks, mobile nodes can switch their interfaces to the ad-hoc mode and form an ad-hoc network. The basic idea of H-WLAN is to enable mobile nodes to dynamically switch between the infrastructure mode and the ad-hoc mode so that they can download objects from the info-station, or cooperatively share their objects with neighboring ad-hoc

nodes. Because these self-organized ad-hoc networks can be assigned the idle channels that are not used by APs, ad-hoc nodes can exploit extra channel bandwidth to disseminate objects and do not need to contend for channel resources with the nodes in the infrastructure mode. Hence, through integrating the infrastructure mode with the ad-hoc mode, H-WLAN can utilize more channel resources to disseminate content, and, hence, achieve higher aggregate throughput.

When designing H-WLAN, we must consider the issue of channel assignment so that the system can balance the load of disseminating objects in the ad-hoc mode across multiple channels. Not surprisingly, an efficient channel allocation scheme is one of the most important factors that determine the performance of multi-channel H-WLAN. In a hotspot, the traffic load for distributing different objects is usually skewed and closely correlated to the popularity of the objects, i.e., the number of requests for the objects. Hence, the system should allocate more channel resources to the popular objects, which are frequently requested and generate heavier traffic load. For example, if there are only two idle channels in an info-station system, the system could allocate one channel for 10 popular objects while allocate the other channel for other non-popular objects. In this paper, we propose a popularity-based channel allocation algorithm that can assign channels to each object based on the popularity of that object. Therefore, as a mobile node requests an object, it can switch to the channel assigned to that object and download the object from one of APs or its ad-hoc neighbors. A requirement before developing a popularity-based allocation algorithm is to identify the popularity of objects. Thus, we also present a rated-based algorithm to estimate the popularity of each object.

In H-WLAN, when a mobile node sends an object downloading request to the info-station, the info-station replies a message that includes the identities of the channels assigned to the requested object. Thus, the mobile node can download the requested object either from the info-station or from the ad-hoc network using the assigned channel. However, the requested object may be replicated in multiple info-stations or ad-hoc nodes, called source candidates. In order to keep load balancing among available sources, the requesting node must determine a proper source node that can provide a better dissemination service, while avoid overload-

ing any of the nodes. Thus, the other goal of H-WLAN is to develop an adaptive source selection mechanism that enables each user to select the best source node. Based on the adaptive source selection mechanism, the requesting nodes can estimate the workload of each source candidate and transmission quality of each communication path, and, hence, can locate the best source to retrieve the requested object in a distributed manner.

In summary, the main contributions of HWLAN are two-fold. First, we design an adaptive source selection mechanism, which can balance the load of cooperative peer-to-peer object dissemination among available source nodes. We carried out an initial study of adaptive source selection in our previous paper [8], and the results of performance evaluation in [8] show that the cooperative integrated-mode hybrid-WLAN system can achieve higher throughput of object dissemination. Second, in this work, we further present a popularity-based channel allocation algorithm to maximize channel utilization and keep the traffic load, i.e., object requests, balancing across channels, i.e., ad-hoc networks. We use simulations to evaluate the performance of two components in H-WLAN. Simulation results show that (1) in H-WLAN with three extra non-overlapping channels allocated to ad-hoc networks, the throughput of object downloading increases by more than 45% in comparison to that in the traditional infrastructure-based info-station; (2) the popularity-based channel allocation algorithm provides 40% higher throughput than naive round-robin channel allocation; (3) when the system has a heavy traffic load, the adaptive source selection mechanism outperforms the random selection scheme, where mobile nodes select source nodes randomly.

The remainder of the paper is organized as follows. In Section 2, we discuss several related works on info-station systems and multi-channel environments. Section 3 presents an overview of the H-WLAN framework, and describes its two major components: *Popularity-based channel allocation* and *adaptive source selection*. In Section 4, we evaluate the performance of the proposed integrated-mode H-WLAN. We then present our conclusions in Section 5.

2. Related work

Most works on ad-hoc networks focus on providing routing algorithms to improve the efficiency

of data distribution applications. In [9], the authors propose two schemes, CacheData and CachePath, in which mobile nodes cache popular data or data paths for other nodes in the same ad-hoc network. In contrast, H-WLAN attempts to improve data access in an infrastructure-based WLAN environment with the help of multiple ad-hoc networks that use orthogonal channels. Other works on infrastructure-based networks try to improve content dissemination and provide better QoS. For example, in the WICAT system [10], each mobile node has a list of objects of interest located at different info-stations. However, the mobile nodes can only download those objects if they are close to the info-station that can provide the objects required to meet the nodes' requests. In [11], the authors propose a framework that supports a "map-on-the-go" application, whereby each info-station provides relevant maps for users moving around the info-station. The systems in [10,11] focus on providing an efficient infrastructure-based wireless system to distribute information, whereas our goal is to design a framework that integrates the infrastructure mode with several ad-hoc networks to cooperatively provide an information dissemination service.

Other approaches, such as [12–14], NUMI [15], and M^2 -WLAN [16], use peer-to-peer communications to assist an infrastructure-based system with data exchange or distribution. In [14], the authors improve the efficiency of content distribution in an info-station system through a cooperative mechanism, i.e., peer-to-peer communication. Two mobile nodes check each other's content lists, identify the objects they need, and exchange the content in a peer-to-peer manner. In NUMI [15], APs collaborate to predict the location and upcoming mobility of mobile nodes. Based on the estimation of a node's movement, an AP can select another mobile node that is likely to move around the vicinity of the requesting node to piggyback data for the requesting mobile node. However, predicting the movements of mobile nodes has a severe impact on the performance of the NUMI system. In the M^2 -WLAN [16], in addition to communicating with the AP, mobile nodes can also switch to the ad-hoc mode to communicate with each other under the administration of the AP, the central manager that assigns channels and stores information about mobile nodes. Both the NUMI and M^2 -WLAN frameworks are designed for host-centric applications, while the proposed H-WLAN framework is

designed for data dissemination applications. In addition, peer-to-peer communication in NUMI and M^2 -WLAN is based on single-hop connection, but it can be a multi-hop connection in the proposed H-WLAN. In other words, the transmission coverage of our framework can be extended through multi-hop forwarding.

Recently, many works, [17–22], have enhanced their respective MAC and routing protocols by exploiting multiple orthogonal channels to provide broadband access in wireless mesh networks. Mesh routers are usually equipped with multiple interfaces and act like fixed routers to relay packets for end devices. In contrast, we focus on how to utilize multiple channels to resolve the load sharing problem experienced by data dissemination applications in some hotspots or info-station systems. In these circumstances, mobile nodes may only have a single interface and are more likely to join networks dynamically to download popular local information, such as lecture slides, travel information, and on-line coupons, from the info-station.

3. The framework of hybrid-WLAN (H-WLAN)

The IEEE 802.11 standard provides multiple non-overlapping channels that can be used simultaneously without interference. In a traditional wireless network, each mobile node can only be a member of a BSS (basic service set) or an IBSS (independent basic service set), and must use a single assigned channel to communicate with mobile nodes in the same BSS/IBSS. A number of previous studies, such as [23,24], noted that mobile devices can access multiple channels by dynamically switching between different channels or alternative operation modes, i.e., the infrastructure mode or the ad-hoc mode. Therefore, the goal of our H-WLAN is to utilize multiple non-overlapping channels to form several cooperative ad-hoc networks to utilize bandwidth capacity of available orthogonal channels and improve the throughput of data dissemination.

An H-WLAN is composed of several info-stations (APs) and ad-hoc networks that provide the cooperative content dissemination service for mobile nodes. The mobile nodes form ad-hoc networks dynamically to help with the distribution of popular objects. In H-WLAN, APs assist each other and act as cooperative controllers to manage these

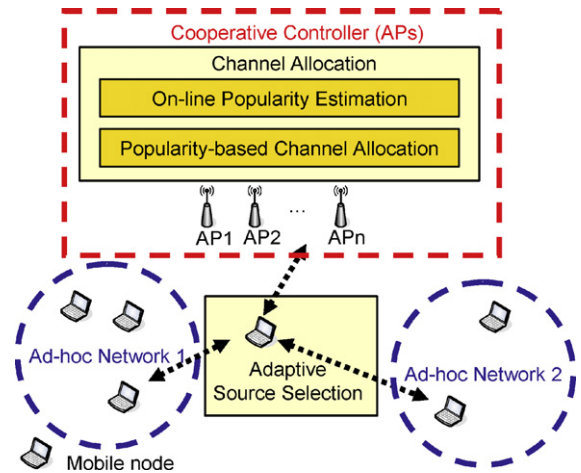


Fig. 1. Framework of H-WLAN. H-WLAN contains two components: (1) *Popularity-based channel assignment*, which is implemented in the info-station to manage channel resources, and (2) *Adaptive source selection*, which enables mobile nodes to select a proper source node in a distributed manner.

co-existing ad-hoc networks. The H-WLAN framework, shown in Fig. 1, contains two major components: *Popularity-based channel allocation* and *adaptive source selection*. In our design, all APs are designated as cooperative controllers to manage channels based on the *Popularity-based channel allocation* algorithm, while the *adaptive source selection* scheme is implemented on each mobile node to select source nodes in a distributed manner. We describe the components in detail in the following.

3.1. Popularity-based channel allocation

To improve multi-channel utilization and balance the load of data dissemination, H-WLAN allocates channels to the objects in proportion to their popularity. In this section, we first present a mechanism that estimates the popularity of objects on-line, and then describe our popularity-based channel allocation algorithm.

On-line popularity estimation: The system can accumulate a number of requests in a given time interval to estimate the popularity of a specific object. However, the accuracy of the estimation depends to a large extent on the length of the sampling interval. A short interval reduces the accuracy because the sample size is too small, while a long interval cannot fully reflect the up-to-date popularity of the object. To avoid having to determine the

Table 1
Definition of symbols

Symbol	Definition
O	the set of all objects
λ_i	the arrival rate of requests for the object i
p_i	the estimated popularity of the object i
u_i	the amount of channel resources utilized by the object i
CH_i	the set of channels that are allocated to the object i
N_{ch}	the number of non-overlapping channels
b_k	the residual available bandwidth of the channel k
$u_{i,k}$	the amount of bandwidth of channel k allocated to the object i

Define parameters used in the popularity-based channel allocation algorithm.

length of an interval, H-WLAN uses a rate-based popularity estimation algorithm, shown in [Algorithm 1](#), to dynamically estimate the popularity of an object. [Table 1](#) summarizes the symbols used in the popularity-based channel allocation algorithm.

Algorithm 1. Rate-based Popularity Estimation Algorithm

```

1: Initial setting:
2: for first request for object  $i$  do
3:    $\lambda_i^{(1)} = 0$ 
4:    $p_i = 1/|O|$ 
5:    $t_i^{(1)} = clock$ 
6: end for

1: On-line estimation:
2: for  $k$ th request for object  $i$  do
3:    $\tau_i^{(k)} = t_i^{(k)} - t_i^{(k-1)}$ 
4:    $\lambda_i^{(k)} = 1/\tau_i^{(k)}$ 
5:    $\lambda_i^{est} = \bar{\lambda}_i = \alpha\bar{\lambda}_i + (1 - \alpha)\lambda_i^{(k)}$ 
6:    $t_i^{(k)} = clock$ 
7:   for all other objects  $j \in O \setminus \{i\}$  do
8:      $\tau_j^{pseudo} = clock - t_j^{last}$ 
9:      $\lambda_j^{pseudo} = 1/\tau_j^{pseudo}$ 
10:    if  $\lambda_j^{pseudo} < \bar{\lambda}_j$  then
11:       $\lambda_j^{est} = \alpha\bar{\lambda}_j + (1 - \alpha)\lambda_j^{pseudo}$ 
12:    else
13:       $\lambda_j^{est} = \bar{\lambda}_j$ 
14:    end if
15:  end for
16: for all objects  $j \in O$  do
17:    $p_j = \lambda_j^{est} / \sum_{obj \in O} \lambda_{obj}^{est}$ 
18: end for
19: end for

```

The popularity of an object i correlates closely with the interarrival time of two successive requests. Assume O is the set of objects in the system. The interarrival time of the k th request for an object $i \in O$ is set as $\tau_i^{(k)} = t_i^{(k)} - t_i^{(k-1)}$, where $t_i^{(k)}$ and $t_i^{(k-1)}$ are the time of the current and previous requests, respectively. Therefore, a shorter interarrival time τ_i indicates that the object i is queried frequently and is therefore more popular. Hence, in H-WLAN, we use the request rate $\lambda_i^{(k)}$ of the k th request (the inverse of $\tau_i^{(k)}$) to represent the popularity of the requested object i . However, the rate calculated based on a single interarrival time cannot represent the average trend, since a temporary traffic load or some exceptional cases are likely to cause prediction errors or system oscillations. Therefore, we must smoothly update the estimate of $\lambda_i^{(k)}$ for each request. This can be achieved by using the following exponential weighted moving average $\bar{\lambda}_i^{(k)}$:

$$\bar{\lambda}_i = \alpha\bar{\lambda}_i + (1 - \alpha)\lambda_i^{(k)},$$

for each object $i \in O$ and $k > 1$, (1)

where α is the filter constant and $0 \leq \alpha \leq 1$. We evaluated the value of α , and found that the system could avoid oscillations and achieve a better performance when α is set to 0.7 or 0.9. The evaluation results of α selection are presented in [Section 4.3](#).

One might notice that the estimate of the request rate of an object in [Eq. \(1\)](#) is only updated when the object is requested. However, if an object has not been requested for a long period, the system should re-estimate (decrease) its request rate and popularity. Hence, whenever the system receives a request for the object i , except estimating the popularity of the object i , the system also needs to update the popularity estimation for all other objects $j \in O \setminus \{i\}$. More specifically, as shown in [Fig. 2](#), for each $j \in O \setminus \{i\}$, the system checks the pseudo request rate of the object j by $\frac{1}{clock - t_j^{last}}$, where t_j^{last} is the time of the last request for the object j and $clock$ is the current system time. Note that the true request rate is always less than the pseudo request rate λ_j^{pseudo} ($\lambda_j^{next} < \lambda_j^{pseudo}$), since the time of the next request for the object j must be larger than clock. Therefore, if the pseudo rate is less than the average rate, it implies that the true request rate of the object j is decreasing, i.e., $\lambda_j^{next} < \bar{\lambda}_j$. Hence, when $\lambda_j^{pseudo} < \bar{\lambda}_j$, the system should recompute the estimate of the request rate λ_j^{est} such that the system can use the estimated request rate λ_j^{est} to represent

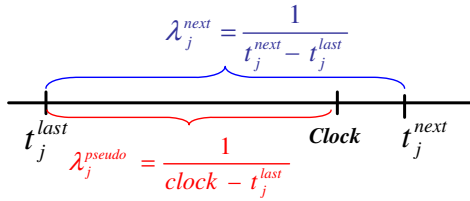


Fig. 2. Rate estimation. Clock is the system time while updating the estimate of popularity. The true request rate of object j (measured according to the upcoming request at t_j^{next}) is always less than the pseudo request rate λ_j^{pseudo} ($\lambda_j^{next} < \lambda_j^{pseudo}$). Therefore, when $\lambda_j^{next} < \lambda_j^{pseudo} < \bar{\lambda}_j$, the system should update $\lambda_j^{est} = \alpha \bar{\lambda}_j + (1 - \alpha) \lambda_j^{pseudo}$ because the rate is decreasing.

the decreasing popularity of the object j . Finally, as shown in line 17 of Algorithm 1, we denote the normalized request rate λ_j^{est} to the popularity of an object j , i.e., p_j , for each object j in the system. Note that the temporal usage of the pseudo rate λ_j^{pseudo} does not affect (change) the real estimate of the request rate $\bar{\lambda}_j$; namely, $\bar{\lambda}_j$ will only be updated until the next request for the object j arrives (i.e., t_j^{next}).

Algorithm 2. Channel assignment algorithm

Require: initially set $u_i = 0$ and normalize $b_j = 1/N_{ch}$

- 1: **while** a node requests to form a group **do**
- 2: Call Channel Deallocation Algorithm
- 3: $b_k = \max_{j=1}^{N_{ch}} b_j$
- 4: **for** each object i in this node **do**
- 5: **if** ($u_i \leq p_i$) and ($b_k > 0$) **then**
- 6: Assign channel k for object i such that $CH_i = CH_i \cap \{k\}$
- 7: **if** ($p_i - u_i$) $\geq b_k$ **then**
- 8: $u_i = u_i + b_k$
- 9: $u_{i,k} = u_{i,k} + b_k$
- 10: $b_k = 0$
- 11: **else if** $0 < (p_i - u_i) < b_k$ **then**
- 12: $b_k = b_k - (p_i - u_i)$
- 13: $u_{i,k} = u_{i,k} + (p_i - u_i)$
- 14: $u_i = p_i$
- 15: **end if**
- 16: **else**
- 17: reject
- 18: **end if**
- 19: **end for**
- 20: **end while**

Popularity-based channel allocation: When a mobile node receives requested objects that it is willing to share, it can ask the cooperative controller to

form an ad-hoc network. The controller then decides whether it is necessary to assign an idle channel to the requesting node so that it can help disseminate its objects in an ad-hoc network. If the cooperative controller determines to allocate the channel resources to the objects held by the requesting node, it will allow the node to switch to the assigned channel and help with the data delivery in the ad-hoc network. Since the popularity of the objects, i.e., the number of object requests, may be skewed, the load of disseminating popular objects must be higher than that of disseminating non-popular objects. Thus, channel assignment should take the popularity into account to avoid a load imbalance across orthogonal channels. In H-WLAN, we employ a popularity-based channel allocation scheme that allocates available channel resources to the objects in proportion to their popularity. The goal of the popularity-based allocation mechanism is to schedule requests of object downloading to proper channels so that the traffic load generated by disseminating various objects can be balanced across multiple channels.

Algorithm 2 is the proposed popularity-based algorithm. Assume there are N_{ch} non-overlapping channels in the system. For each object i , the amount of channel resources allocated to it is initially set as $u_i = 0$. We denote b_j as the residual bandwidth of the wireless channel j . Each b_j for $j = 1, \dots, N_{ch}$ is initially set as $1/N_{ch}$ so that the bandwidth of the total channel resources is normalized to 1, which can be fully utilized by our channel allocation algorithm. When a mobile node requests to switch to the ad-hoc mode, the controller determines if it is necessary to assign channel resources to the objects provided by that node. Because we normalize the total bandwidth resources (b_j) and the overall popularity of objects (p_i), to 1, the popularity-based algorithm can allocate a proportion of channel resources to each object based on its normalized popularity. That is, if u_i , the normalized channel resources allocated to the object i , is equal to p_i , the normalized estimated popularity of that object, it means the object i held by the requesting node has already utilized (allocated) its proportional channel resources, i.e., $(p_i * N_{ch})$ channels. Therefore, the system does not assign additional channel resources to the node's request, and only permits it to join the existing ad-hoc network and assist in distributing the objects over the channel that has been assigned to its objects previously.

Algorithm 3. Channel deallocation algorithm

```

1: for each object  $i \in O$ 
2:   if  $u_i > p_i$  then
3:     for each channel  $k \in CH_i$  do
4:        $b_k = b_k + \frac{u_{i,k}}{u_i}(u_i - p_i)$ 
5:        $u_{i,k} = u_{i,k} - \frac{u_{i,k}}{u_i}(u_i - p_i)$ 
6:     end for
7:      $u_i = p_i$ 
8:   end if
9: end for
10: for each object  $i$  that is removed from the system do
11:   for each channel  $k \in CH_i$  do
12:      $b_k = b_k + u_{i,k}$ 
13:      $u_{i,k} = 0$ 
14:   end for
15:    $u_i = 0$ 
16:    $CH_i = \emptyset$ 
17: end for

```

Otherwise, if the utilized channel resources u_i is lower than the estimated popularity p_i for at least one of its objects, the controller assigns the channel with maximal residual bandwidth, say channel k , to it. That is, the requesting node holding the object i can join the ad-hoc network using the channel k^1 to provide a content dissemination service. For each object i with $u_i < p_i$, the amount of channel resources that could be allocated to it is $\min(p_i - u_i, b_k)$ (the $\min(a, b)$ function returns the smaller of two input values). Then, the residual bandwidth resources b_k for channel k is reduced by $\min(p_i - u_i, b_k)$, and the amount of allocated resources u_i for the object i is increased by $\min(p_i - u_i, b_k)$. After each round of allocation, if there is still residual bandwidth in the channel k , i.e., $b_k > 0$, this means the system can allocate the channel k to other objects. Similarly, if the amount of channel resources allocated to the object i , i.e., u_i , is still smaller than the popularity p_i , the system can assign other idle channels to other ad-hoc nodes that also hold the object i in the following rounds of allocation. Hence, in our system design, each object is allocated the amount of channel resources that is proportionate to its popularity.

In our channel allocation scheme, it seems that bandwidth of a channel can be partitioned into several parts; however, we use the idea of sharing the channel resources to represent that the ad-hoc nodes

holding less popular objects should use the same channel because of their light traffic load. However, the ad-hoc nodes holding the popular objects can utilize more channel resources to provide dissemination service and balance the load across channels.

The objects assigned some bandwidth resources may become unpopular after allocation; hence, the system needs a channel deallocation scheme that can retrieve the unutilized resources such that available bandwidth can be allocated to up-to-date popular objects. Instead of periodical deallocation, the H-WLAN executes the deallocation algorithm before resource allocation for each node request, as shown in line 2 of Algorithm 2. Algorithm 3 presents the proposed deallocation algorithm. In the deallocation algorithm, when the amount of channel resource allocated to the object i , u_i , is higher than its up-to-date popularity estimation p_i , the system retrieves the amount of channel resource $(u_i - p_i)$ and resets the utilization value $u_i = p_i$. However, the system may assign multiple channels to an object, so it needs to know which channels are utilized by the object i and retrieve the bandwidth resources of those channels. Hence, in Algorithm 2, the system maintains a set of channels, CH_i , that collects the channels allocated to the object i , and records the amount of bandwidth of the channel k that is allocated to the object i , denoted by $u_{i,k}$, for each channel $k \in CH_i$. When the system detects that the channel resources utilized by the object i should be deallocated, i.e., $u_i > p_i$, it proportionally retrieves bandwidth $\frac{u_{i,k}}{u_i}(u_i - p_i)$ from each channel $k \in CH_i$. For example, assume an object i is allocated 50% resource, i.e., $u_i = 0.5$, where 30% from the channel 1 ($u_{i,1} = 0.3$) and 20% from the channel 2 ($u_{i,2} = 0.2$). If the popularity of the object i becomes 0.2, the system retrieves 30% resources back: $\frac{u_{i,1}}{u_i}(u_i - p_i) = \frac{0.3}{0.5}(0.5 - 0.2)$ from channel 1 and $\frac{u_{i,2}}{u_i}(u_i - p_i) = \frac{0.2}{0.5}(0.5 - 0.2)$ from channel 2. On the other hand, if an object i is removed from the system because all nodes that can provide this object leave, the system retrieves all resources assigned to it, i.e., u_i , and resets CH_i to an empty set as shown in line 11 to line 16 of Algorithm 3.

3.2. Adaptive source selection

In H-WLAN, an object may be replicated in multiple info-stations or ad-hoc nodes, called source candidates. Thus, when the cooperative controller receives an object request from a mobile node, it responds a source candidate list, i.e., a list of source

¹ We assume that each node is equipped with one interface, and can only assigned a single channel.

Table 2
Group-information table

Ad-hoc network ID	Channel assignment	Member ID	Object list
1	1	6	2, 5, 6
		8	1, 3
		(Gateway) 9	2, 6
2	6	4	3
		7	3, 4, 8
		(Gateway)	

The controller records the status of ad-hoc networks, i.e., the mobile nodes in each ad-hoc network and the object list of each node.

candidates, to the requesting node. In this paper, we propose an adaptive source selection scheme, which can direct mobile nodes to select the best source from the source candidate list and, at the same time, balance the load among various source candidates. In the following, we first describe how the controller manages the ad-hoc networks and compiles a source candidate list for the requesting node, and then present the proposed adaptive source selection scheme.

Ad-hoc network management: In a H-WLAN, the cooperative controller maintains a group-information table, shown in Table 2, to record the status of each ad-hoc network, i.e., the mobile nodes in each ad-hoc network and the object list of each node. In Table 2, the member ID field records the identity of each ad-hoc node, while the object list field keeps indices of the objects held by the corresponding ad-hoc node. The channel assignment field maintains the channel allocated to that ad-hoc network. While the controller assigns an unused idle channel to form an new ad-hoc network, the mobile node initiating the ad-hoc network is designated as the gateway, which should periodically switch back to the infrastructure mode to update the group-information to the controller. The controller needs to dynamically update the group-information table when the status of the ad-hoc networks changes, i.e., nodes leave or join the ad-hoc networks under the following conditions.

- *To retrieve an object:* If a mobile node selects an ad-hoc node as its object source, it should first join the same ad-hoc network as the source node to download the requested object.
- *To share an object:* When a mobile node offers to share its objects, but is not allocated a new channel, it can join an existing ad-hoc network.

- *Outside an AP's coverage range:* A mobile node that can not connect with any AP can join one of ad-hoc networks to obtain desired objects. The mobile node locates an ad-hoc network by broadcasting beacon messages; therefore, location information is not required in H-WLAN.
- *Switching to another BSS/IBSS:* When an ad-hoc node can not get the required objects from any nearby ad-hoc nodes, it can switch to other ad-hoc networks or switch back to the infrastructure mode to get the objects.
- *Leaving the network:* The mobile node should inform the cooperative controller before it leaves the network so that the controller can update the group-information table. If the departing node is the gateway of an ad-hoc network, it should appoint another node that is in the same ad-hoc network and can communicate with the controller directly as the new gateway.

Because the mobile nodes join or leave ad-hoc networks dynamically, some objects that have been allocated the channel resources may not exist in ad-hoc networks anymore. In that case, the controller needs to retrieve the channel resources and recompute the allocated channel resources u_i for those objects i .

The cooperative controller receiving the request checks the group-information table (Table 2) and generates a source candidate list, as shown in Table 3, for the requesting node. In the source candidate list, each tuple of object i contains two fields: one is the wireless channel assigned to this object, while the other identifies the ad-hoc nodes or APs holding the requested object. Next, we describe how a mobile node, depending on its location, obtains the source candidate list.

Table 3
Source candidate list

Object ID	Channel ID	Available source ID
1	1	8
	11	AP1
2	1	6, 9
	11	AP1
3	1	8
	6	4, 7
	11	AP1

Each tuple contains the wireless channel assigned to the requested object and the set of available source nodes.

- *Within the AP's coverage range:* If the mobile node is in the infrastructure mode, it can request the source candidate list from the cooperative controller directly. However, if it operates in the ad-hoc mode, it has to switch back to the infrastructure mode, without affecting ongoing communications, to get the source list. That is, the mobile node has to finish the current transmission in the ad-hoc network and then switch back to the infrastructure mode.
- *Outside the AP's coverage range:* When the mobile node is outside the transmission range of any AP, it can join any ad-hoc network within its direct transmission range and then forward the request to the gateway of that ad-hoc network. Because the gateway periodically switches back to the infrastructure mode to exchange information with the cooperative controller, it can query the source candidate list for the requesting node. Therefore, even though some mobile nodes cannot connect with any AP directly, they still have the opportunity to get the requested objects via cooperative content sharing in an ad-hoc manner.

Adaptive source selection: A mobile node can choose an appropriate source node from the source candidate list to get desired objects. The objectives of the adaptive source selection scheme are to improve transmission quality of data dissemination and balance the load among all source nodes. In this work, we propose a ping-based source selection method that enables mobile nodes to measure transmission quality by using probe messages. When a mobile node obtains the source candidate list, it can switch between different channels to ping each source candidate and evaluate quality of each communication path. The ping-based scheme uses the response time of a short probe message to examine the round-trip delay and loss probability of the communication path between the requesting node and the source. The round-trip delay includes the time spent for transmitting a probe message, the waiting time experienced by the response message in the source's transmission queue, and the time spent for receiving the response from the source. The transmission time identifies the traffic load in wireless medium, while the queueing delay can indicate the workload of the source node. The loss probability reflects the channel quality of an wireless link between two mobile nodes. If the requesting node receives a reply from the source

quickly and with low loss probability, it means the node can download the object from the lightly loaded source through a good communication path.

However, when there are too many source candidates, the ping-based method may consume too much time and channel resources for collecting ping information. In addition, when multiple nodes probe ping messages at the same time, they may select the same source node and overload that node. Hence, we consider two variations of the ping-based method, the ping-random-2 (PR2) method and the random-2-ping (R2P) method, to resolve the above problems. The PR2 method is similar to the ping-based method, except that the mobile node has to choose a source randomly from the two source candidates that return their response messages most quickly. In the R2P method, on the other hand, the mobile node randomly selects two nodes from the source candidate list and only probes those two nodes, instead of all the candidates on the list. In the PR2 and R2P methods, since mobile nodes can incorporate random technique into ping-based measurement, it is less likely to connect to the same source as other mobile nodes. Therefore, two variations of the ping-based method can select the best source via investigating transmission quality, while avoid overloading any of the data sources. In addition, the R2P method typically reduces the probing overhead.

4. Performance evaluation

In this section, we use the terminating simulation [25,26] to evaluate the performance of H-WLAN within the whole period of its operation, which includes the formation of ad-hoc networks, the converge period, and the termination of ad-hoc networks. We describe the simulation scenarios and metrics used to evaluate the performance of H-WLAN, and then show the results of the comparison between the baseline methods and H-WLAN. The simulations are implemented in an ns-2.28 environment with a CMU wireless extension [27], and run on Linux-2.6.20 machines. For each scenario, we repeat simulations for 20 iterations, and vary the seed of NS2 pseudo random number generator (PRNG) from 1 to 20 to generate the random variables. All figures present the average of 20 iteration results, and report the 95% confidence interval.

4.1. Simulation environment and performance metrics

In the simulations, three APs are uniformly distributed in a $450 \text{ m} \times 400 \text{ m}$ area. The capacity and delay of wired links connected to APs are set as 100 Mbps and 2 ms, respectively. Six 802.11a non-overlapping channels are utilized in the simulations, and the base rate of each wireless channel is 6 Mbps. Three APs are assigned three non-overlapping channels, and the other three channels are available for setting up three ad-hoc networks to facilitate peer-to-peer content sharing. The propagation model used in the simulations is the two-ray ground reflection model. The transmission range of all mobile nodes, including the APs, is 250 m. We assume that the mobile nodes are static as they are downloading objects. The simulation settings are summarized in Table 4. Other parameters that are not listed in Table 4 are set to NS2 default settings. Reference [28] gives a detailed description

about physical layer, data link layer, and IEEE 802.11 MAC protocol used in NS2.

$$\begin{aligned} \text{Prob}[\text{request of object } i] &= \frac{c}{i^{(1-\theta)}} \quad \forall i \\ &= 1, 2, \dots, K \text{ and } 0 \leq \theta \leq 1, \text{ where } c \\ &= \frac{1}{H_K^{(1-\theta)}} \text{ and } H_K^{(1-\theta)} = \sum_{j=1}^K \frac{1}{j^{(1-\theta)}}. \end{aligned} \quad (2)$$

Table 5 summarizes the default scenario settings. The total number of objects is 100, and the size of the objects is uniformly distributed between 500 KB and 5 MB. There are 50 mobile nodes uniformly distributed in the simulation area, and each mobile node requests 10 objects at random. For all default scenarios, we simulate two types of object request patterns, i.e., the distributions of object popularity: the uniform distribution and the zipf distribution. The uniform distribution of popularity means that all objects have an equal probability to be requested

Table 4
Simulation setting

Variable	Setting
MAC	IEEE 802.11a with base rate 6 Mbps
Propagation model	TwoRayGround
Antenna	Omnidirectional antenna
Ad-hoc routing protocol	DSDV
Transmission range	250 m
Content transmission protocol	FTP
Capacity of wired links connected to APs	100 Mbps
Delay of wired links connected to APs	2 ms
Queue type of wired links connected to APs	DropTail
Mobility model	Static (Assume that nodes do not move when they are downloading objects)

Table 5
Default scenario setting

Variable	Setting
Simulation area	$450 \text{ m} \times 400 \text{ m}$
Number of access points	3
Number of channels for APs	3
Number of channels for ad-hoc nodes	3
Number of nodes	50
Distribution of node location	Uniform
Number of total objects	100
Number of requests per user	10
Distribution of object popularity	(1) Uniform, (2) Zipf
Distribution of object size	Uniform distribution between 5 MB and 50 MB
Interarrival time of two-phase arrival	Exponential distribution with mean 1 and 6(/min)
Request interarrival time	Exponential distribution with mean 2(/min)
Channel allocation algorithm	Random scheme

by mobile nodes. On the other hand, the zipf distribution given in Eq. (2) has been shown in [29] to characterize the popularity of web access and content access. Hence, we also evaluate H-WLAN under the zipf distribution of popularity. We set K in Eq. (2) to the total number of objects and $\theta = 0$ to skew the distribution of popularity. In the classic version of Zipf's law, θ is equal to 0, and $\frac{c}{i^{(1-\theta)}}$ is the normalized fraction of time that the object i is requested. For each user, the interarrival time between two requests follows an exponential distribution with mean 2(/min). Another important observation is that users normally request objects at a specific time, such as the time right before a meeting or a class; hence, the interarrival time between two users entering the system gets smaller as the scheduled time of a presentation or class approaches. To evaluate the above scenario, we use a two-phase arrival process in the simulations. The process has a lower arrival rate λ_l for the first k arrivals and a higher arrival rate λ_h for the remaining arrivals. Specifically, the random variables X_l and X_h are used to represent the interarrival times in the first phase and second phase, respectively. The probability distribution function of X_l and X_h is an exponential distribution with means $\lambda_l = 1$ and $\lambda_h = 6$ (/min). In the following simulations, first 5 arrivals are in the phase of the lower rate. The other specific settings for various scenarios are described in the following sections.

The performance metrics used in our work are: (a) the throughput of object downloading, i.e., the size of the requested object divided by the response time required to select the source node, collect ping information, and retrieve content; (b) the control overhead: the ratio of the load generated by probing and exchanging control messages to the total traffic load (including object downloading and control message exchange). Next, we discuss the performance evaluations of *adaptive source selection* and *popularity-based channel allocation* in H-WLAN.

4.2. Source selection study

To illustrate the benefits of the integrated-mode networks and neutralize the effect of the channel allocation scheme, we explore an environment which uses a naive random channel allocation scheme in source selection study. We implement the following baseline methods for comparison with the proposed ping-based source selection scheme.

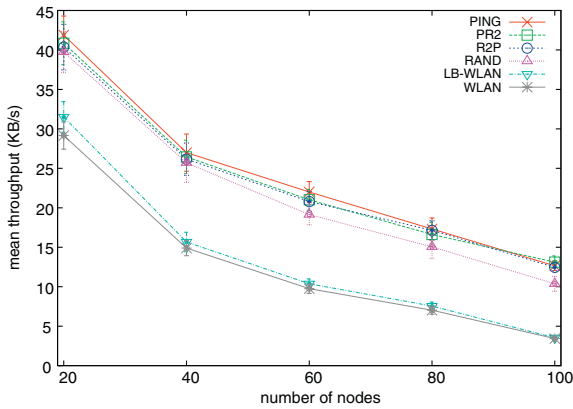
- *Signal-strength method (WLAN)*: WLAN is a conventional info-station system, where mobile nodes connect to the AP with the strongest signal to download the required objects in the infrastructure mode.
- *Log-based method (LB-WLAN)*: In the log-based WLAN, which also operates in the infrastructure mode only, each AP keeps track of the total number of associated traffic flows. A mobile node can connect to a lightly loaded AP to avoid congestion and mitigate the load-imbalance problem.
- *Random selection method (RAND)*: This method is similar to the integrated-mode H-WLAN. However, the mobile node randomly chooses a source from the source candidate list and then checks if it is reachable. If it is not, the node repeats the random selection procedure until it can connect to one of neighboring ad-hoc nodes or APs.

We evaluate the performance of the source selection scheme under several scenarios. For the default scenario presented in Table 5, we examine the impact of the number of clients, the number of total objects, and the number of channels on the performance of H-WLAN. Besides, we consider two additional scenarios, shown in Table 6, to further evaluate the H-WLAN framework under various special network conditions.

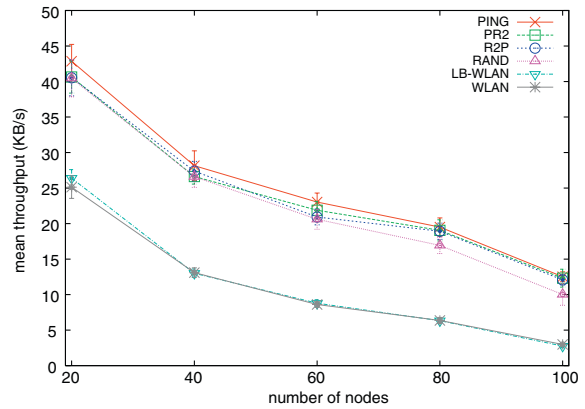
1. *Comparison between integrated-mode and infrastructure-mode*: The simulation results in Fig. 3a and b shows the average throughput of all schemes for various numbers of clients. The figures show that when the mobile nodes are uniformly distributed in the hotspot, LB-WLAN and WLAN have similar performance because the traffic load generated by mobile nodes are distributed equally across APs. Generally, the throughput decrease as there are more clients contending for channel resources with each other. The integrated-mode methods, i.e., PING, PR2, R2P, and RAND, perform more than 45% higher throughput than traditional

Table 6
Special scenarios setting

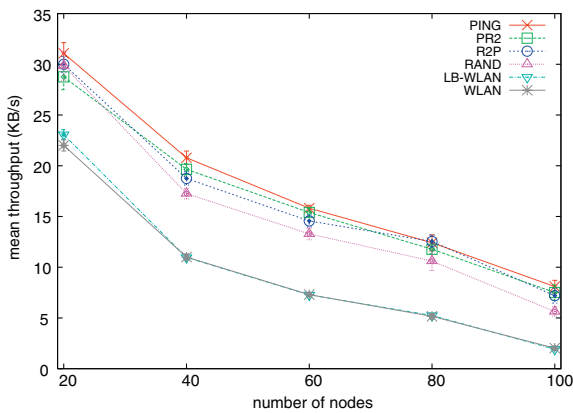
Scenario	Setting
Bulk arrival pattern	Bulk arrival size: 5, 10, 15, 20, 25
Non-uniform location distribution	Percentage of nodes located in the transmission range of a random selected AP's: 50–90%



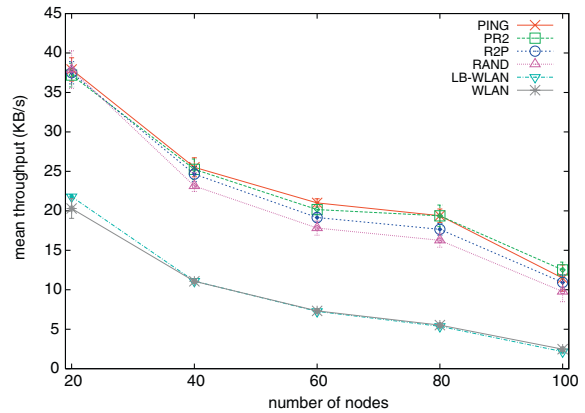
(a) Uniform popularity (two-phase arrival rate)



(b) Zipf Popularity (two-phase arrival rate)



(c) Uniform popularity (constant arrival rate)



(d) Zipf Popularity (constant arrival rate)

Fig. 3. Comparison between Integrated-Mode and Infrastructure Mode. The integrated-mode methods (ping-based and RAND) can benefit from the ad-hoc networks using idle channel capacity. The ping-based methods can locate a proper source and achieve higher throughput than the random scheme.

infrastructure-based systems, i.e., WLAN and LB-WLAN, because integrated-mode networks utilize multiple idle channels to increase both network capacity and aggregate throughput. Although all integrated-mode methods, both ping-based and RAND methods, perform better than infrastructure-based WLANs, we find that the adaptive ping-based methods, such as PING, PR2, and R2P, have even higher throughput than the RAND method, as shown in Fig. 3. This is because, in ping-based methods, a mobile node uses probing mechanism to evaluate each source candidate and then chooses the best source, which operates over a non-congested channel and has light workload. In the RAND method, however, mobile nodes just randomly select a neighboring node as its source. Specifically, in the RAND method, a node could

choose the heavily loaded or farthest source node, which may experience high loss rate over the communication path; thus, it would not get satisfactory QoS. Consequently, the RAND method can not guarantee a steady performance.

In contrast to the two-phase arrival scenario, we also evaluate the performance of H-WLAN under the scenario of constant arrival rate. That is, the interarrival time follows an exponential distribution with a fixed mean $\lambda = 6$ (/min). Fig. 3c and d shows that H-WLAN can benefit from cooperative content sharing over multiple channels in the constant arrival rate scenario as well.

The control overhead in this study consists of the load generated by probe messages and control message exchange between ad-hoc nodes or between the controller and ad-hoc nodes. We let the traditional

infrastructure-based WLAN be the baseline for comparison with other schemes, and do not consider the message exchange in the traditional infrastructure-based WLAN as the overhead of H-WLAN. That is, the WLAN scheme has no extra control overhead. Nevertheless, the LB-WLAN method and the integrated-mode methods require additional channel resources to exchange information, such as source lists and probe messages. As Fig. 4 shows, the overhead of ping-based schemes is no more than 0.4%. Particularly, the R2P methods only probe two source candidates, and reduces the overhead signif-

icantly. The figures show that H-WLAN achieves a significant performance improvement in terms of throughput with relatively low overhead.

2. *Impact of the number of channels on system performance:* We vary the number of channels used to form ad-hoc networks from 2 to 5 to evaluate the benefit of multiple channels. Because the WLAN and LB-WLAN schemes do not use idle channels to form ad-hoc networks, we extend the results of these schemes as straight lines. The results in Fig. 5 show that, in general, the throughput increases when the number of channels increases. When the system assigns more idle channels for

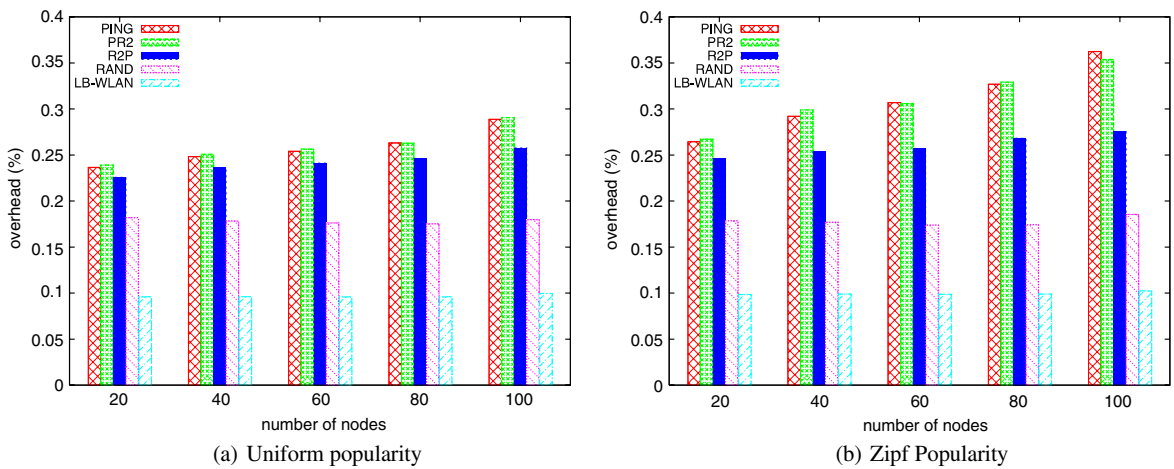


Fig. 4. *Overhead.* The overhead of control messages used to collect ping information and obtain the source list is less than 0.4% of total traffic load.

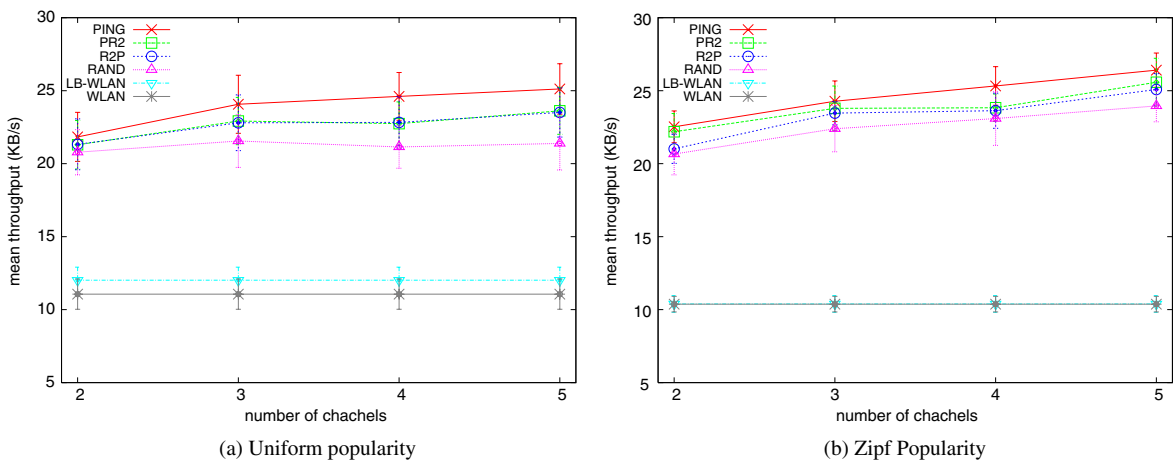


Fig. 5. *Impact of the number of channels.* As the number of channels increases, the ping-based schemes can distribute the load among sources using different channels and, hence, achieve higher throughput improvement in comparison with the random method.

building ad-hoc networks, mobile nodes can utilize more channel resources to exchange their objects and provide higher throughput. In addition, when the number of channels increases, the ping-based schemes can distribute the load among sources using different channels, and, hence, achieve higher throughput improvement in comparison with the random method.

3. *Impact of the number of total objects on system performance:* Fig. 6 shows the average throughput for various numbers of total objects. Fig. 6a shows that the throughput decreases as the number of total objects increases in the uniform popularity scenario, i.e., each object has same popularity. When the number of total objects increases, the probability that different users request the same object decreases. Hence, as the users do not share similar preference for objects and can not exchange the objects in the ad-hoc networks operating over idle channels, the performance improvement of H-WLAN is limited. However, in general, the distribution of content preference follows the zipf distribution, instead of uniform. Hence, Fig. 6b shows that as the popularity is skewed, i.e., users have common preference on some objects, the performance of H-WLAN will not decrease due to the increasing number of total objects.

4. *Bulk arrival pattern:* We now consider which method can perform better under a bursty traffic pattern. In this simulation, all mobile nodes request a single identical object, but the arrival size may be greater than one node. This is called *bulk arrival*. The bulk arrival size denotes the number of requests

that arrive in a very short interval. That is, the arrival size of the first 5 arrivals is 1 and the bulk arrival size of the remaining arrivals is sb , where sb varies from 5 to 25. There are 30 mobile nodes uniformly distributed within the simulation area. As shown in Fig. 7, both the PR2 and R2P methods have higher throughput than the PING method. When the arrival of requests is not bursty, the PING method can choose an appropriate source node based on ping information. However, when the bulk arrival size increases, all users that arrive at the same time may choose the same best source node. During the phase of collecting ping information, the small ping messages may not cause congestion, so all users can still get the response from the best source quickly

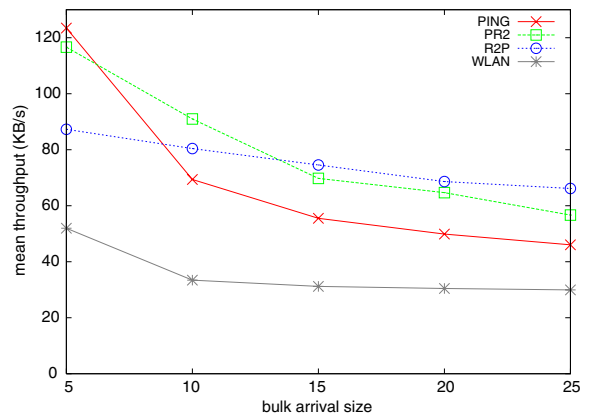
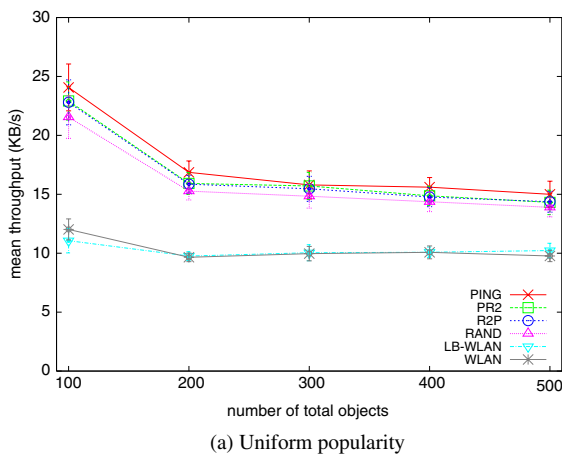
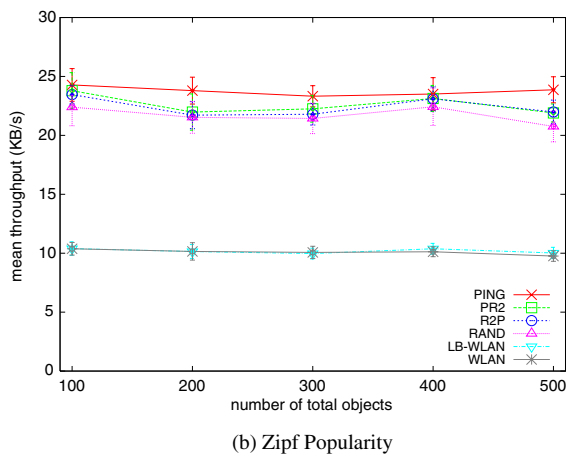


Fig. 7. Bulk arrival pattern. PR2 and R2P can incorporate random technique to avoid overloading any of sources, and, hence, outperform the PING method in the bulk arrival scenario.



(a) Uniform popularity



(b) Zipf Popularity

Fig. 6. Impact of the number of objects. H-WLAN can gain significant throughput improvement as users have similar preference for objects, i.e., the skew-popularity scenario.

such that they may select the same source node. However, when multiple users start to download the object, the heavy traffic load must overload the selected source and lead to congestion in the assigned channel. To resolve the above problem, the PR2 and R2P methods incorporate a random mechanism into the ping-based selection scheme so that mobile nodes can select an appropriate data source by measuring the network's condition and, at the same time, keep the load balancing among all available sources by random technique. This explains why the PR2 and R2P methods perform better than the PING method under the bulk arrival scenario.

5. *Non-uniform location distribution*: There are 35 mobile nodes distributed in a non-uniform manner in the simulation area. All mobile nodes request a single identical object. In Fig. 8, the x -axis denotes the percentage of mobile nodes located in a randomly selected AP's transmission range, but not in the transmission range of other APs. When the traffic pattern in the network is non-uniform, the number of connections going through some APs must be higher than others APs if mobile nodes associate with an AP only based on signal strength. That is, there exists a load-imbalance problem among the APs and among different channels. Even though LB-WLAN addresses the load-imbalance problem among APs, its performance is still limited by the capacity of channels assigned to APs and, hence, below that of the integrated-mode H-WLAN. H-WLAN can provide higher throughput because mobile nodes can obtain objects from ad-hoc net-

works using the idle channels. Thus, they do not have to contend with other users for the channels used by heavily loaded APs. With assistance of ad-hoc networks, the load on the channels allocated to APs can be reduced; thus, congestion in the heavily loaded APs can be relieved. Fig. 8 also shows that H-WLAN achieves a significant throughput improvement when 70% of mobile nodes are distributed in one AP's transmission range.

4.3. Channel allocation study

As noted earlier, the performance of H-WLAN depends on proper source (the routing path and the data source) and efficient channel management. Thus, the motivation of the channel allocation study is to (a) determine the impact of channel allocation on the performance of H-WLAN; (b) tune the filter constant α in the popularity estimation algorithm to optimize the efficiency of the popularity-based channel allocation scheme; and (c) examine how the allocation and deallocation schemes adapt to dynamic popularity change. We use the PING method to evaluate the performance of the channel allocation scheme. Without loss of generality, we let the popularity of objects follow the Zipf distribution to skew preference on different objects. In this simulation, the first- k and round-robin (RR) methods are implemented as base cases and compared with our popularity-based mechanism.

In the first- k method, available channels can only be allocated to first k requested objects. Since a channel in the first- k method can only be allocated to an ad-hoc network to distribute a single object, the system can only build three ad-hoc networks to share at most three objects. In the first- k method, channel resources are not fully utilized because one ad-hoc network using an orthogonal channel only supports dissemination of one object. In the RR method, the cooperative controller assigns channels in a round-robin manner regardless of the utilization of each channel. The mobile node that volunteers to provide an object can join the ad-hoc network that is allocated a channel by the round-robin method. Because objects can be distributed in multiple channels, channel utilization can be improved. However, the load of ad-hoc networks in different channels may differ because the system allocates the channel resources to each object with equal probability; in other words, some channels may be shared by several popular objects, while

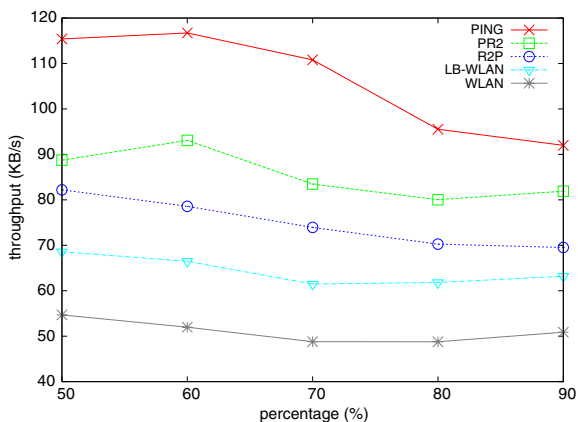


Fig. 8. *Non-uniform location distribution*. Nodes in the high density area can form ad-hoc networks using idle channels and, hence alleviate congestion in the heavy-loaded AP's channel.

others are used inefficiently for disseminating less popular objects.

In Fig. 9a, we vary the number of nodes from 20 to 100 to evaluate the benefit of the popularity-based channel allocation scheme under various densities of network topology. The figure shows that our popularity-based mechanism achieves the best performance in terms of throughput under skewed popularity. The popularity-based channel allocation algorithm performs 15% better than naive round-robin channel allocation when there are 20 mobile nodes in the system, while performs about 300% better than round-robin allocation as there are 100 mobile nodes in the system. The improvement is particularly significant as there is a heavier traffic load in the network; that is, without a proper channel management mechanism, heavy traffic load might causes congestion in some wireless channels, while some other channels may not be utilized efficiently. However, our dynamic popularity-based channel allocation method can schedule multiple flows, i.e., a number of object downloading, to different channels based on the load of each channel; hence, mobile nodes can utilize non-overlapping channels optimally and achieve higher throughput. Fig. 9b shows the performance of H-WLAN for various numbers of idle channels allocated to ad-hoc networks. The gap between the performance of the popularity-based scheme and that of the other two baseline schemes becomes larger as there are more channel resources in the system. It shows that the popularity-based algorithm can directs object requests to orthogonal channels equally and utilize channel

capacity efficiently. Without a good channel allocation mechanism, the system cannot guarantee optimal channel utilization because of the load-imbalanced problem; that is, some channels may be over-congested while others are still idle.

The estimation of the popularity of objects is updated on-line once a request is made. The higher throughput also shows that the estimation mechanism in H-WLAN is accurate enough to optimize the benefits of popularity-based channel allocation. Fig. 10 shows the impact of α on the performance of popularity-based channel allocation. H-WLAN performs better when α is set at 0.7 and 0.9. A small α value causes system vibration, since the estimation may be heavily influenced by a single sample. An

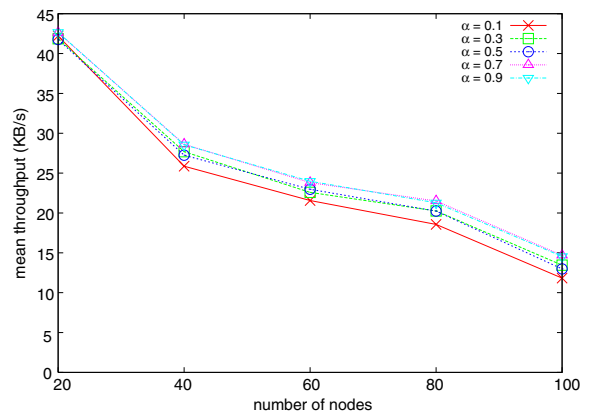
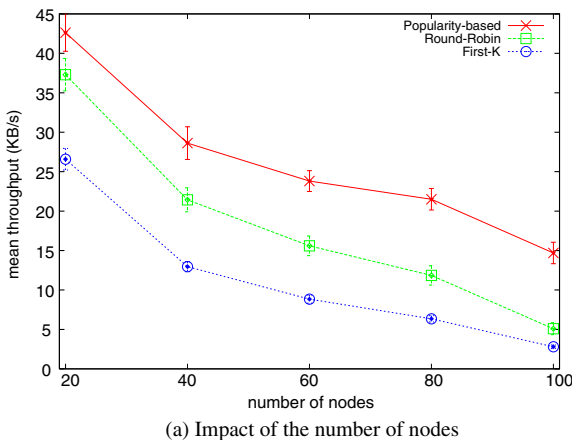
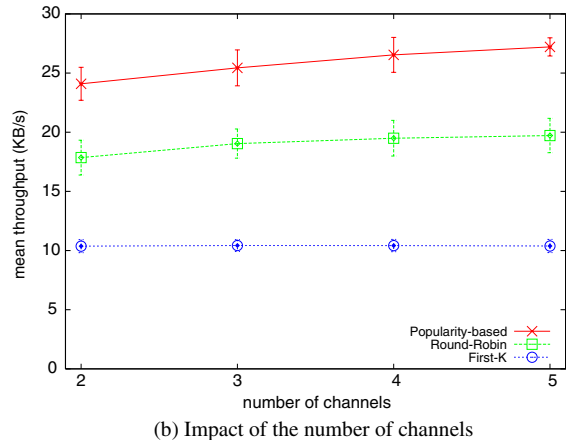


Fig. 10. Effect of α on popularity estimation. When α is set at 0.7 or 0.9, the channel allocation algorithm can provide higher throughput based on more accurate estimation.



(a) Impact of the number of nodes



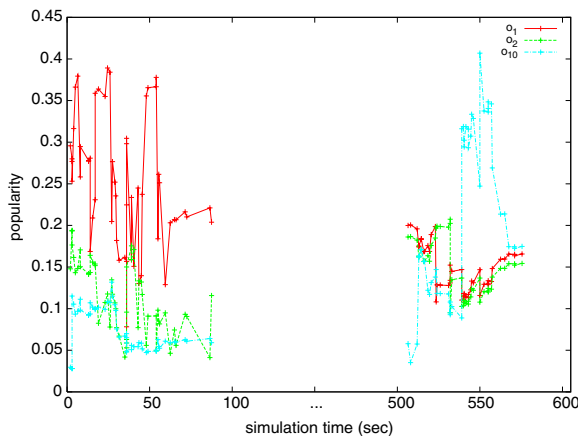
(b) Impact of the number of channels

Fig. 9. Comparison of channel allocation schemes. The performance improvement increases as the system has a heavy traffic load or as there are more channels. It indicates that the popularity-based allocation can efficiently utilize multiple channel capacity to disseminate objects.

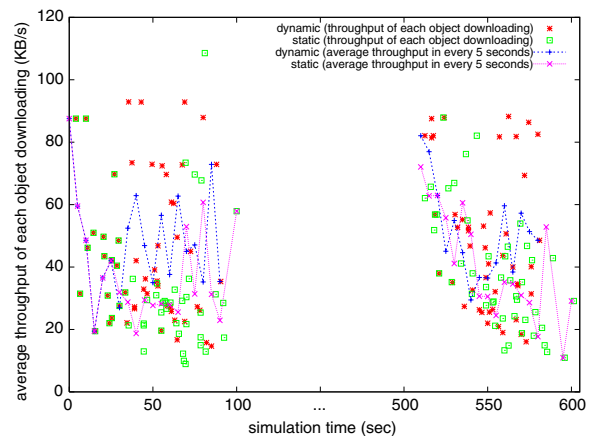
appropriate α value can guarantee accurate estimation and improve the efficiency of channel allocation.

Finally, we evaluate how the rate estimation and channel allocation algorithms adapt to dynamic popularity change. In this simulation, there are 50 mobile nodes in the system, and each of them requests for two objects. There are 10 objects, $O = \{o_1, o_2, \dots, o_{10}\}$, in the system; without loss the generality, the popularity of objects follows the zipf distribution. However, we divide the simulation into two phases: (1) $t < 100$, and (2) $t > 500$, where t is simulation time. In the first phase, let the object 1 be the most popular object, i.e., $p_{o_1} > p_{o_2} > \dots > p_{o_{10}}$. Between $t = 100$ and $t = 500$, no request arrives the system. After the idle period, i.e., $t > 500$, we invert the popularity distribution in the second phase such that the object 10 becomes the most popular object, i.e., $p_{o_{10}} > \dots > p_{o_2} > p_{o_1}$. Fig. 11a shows how the popularity estimation algorithm updates the estimate of popularity p_{o_1}, p_{o_2} , and $p_{o_{10}}$ as the popularity of objects changes. After $t = 500$, the estimated popularity values of the object 1 and 2 decrease, while the popularity of the object 10 increases. Since the proposed estimation algorithm uses λ^{pseudo} to detect popularity decline, the system can decrease the estimated popularity value of the unpopular objects (p_{o_1}) despite no request for those objects.

Fig. 11b compares the performance of the dynamic and static popularity-based channel allocation algorithms in terms of the throughput of each object downloading as well as the average throughput in every 5 s interval. The x -axis denotes the time that a downloading is finished, while the y -axis means the average throughput of that downloading. The static algorithm is similar to the proposed channel allocation algorithm except (1) the system does not deallocate channel resources when the objects become unpopular and cannot utilize the assigned channels efficiently, and (2) the system estimates the popularity only based on the rate estimation shown in Eq. (1) without considering the pseudo request rate λ^{pseudo} . The results in Fig. 11b show that the proposed allocation/deallocation algorithm can perform better than the static scheme when $t > 40$ because the system can retrieve channel resources and reallocate them to other popular objects. In addition, in the static scheme, the throughput of each object downloading decreases when popularity changes, i.e., when $t > 550$. However, the proposed real-time popularity adaptation and resource deallocation schemes can guarantee stable throughput of each object downloading even though the popularity distribution extremely changes. Fig. 11b also shows that the time needed for finishing all downloading in the static scheme (600 s) is longer than the dynamic algorithm (580 s) because the static



(a) Real-time popularity estimation



(b) Throughput of each object downloading

Fig. 11. Performance of the channel allocation algorithm under dynamic scenario. In the first phase ($t < 100$), $p_{o_1} > p_{o_2} > \dots > p_{o_{10}}$. The system is idle during $100 < t < 500$. In the second phase ($t > 500$), the popularity distribution is inverted such that $p_{o_{10}} > \dots > p_{o_2} > p_{o_1}$. Figure (a) only shows the objects that have dynamic popularity change, i.e., object 1, 2, and 10. In Figure (b), the points are the throughput of each object downloading, while the lines illustrate the average throughput in every 5-s interval. The results show that real-time popularity adaptation and resource deallocation can guarantee stable throughput of each object downloading even though the popularity extremely changes.

scheme might starve some nodes who download the popular objects in the second phase, such as the object 10, but cannot fairly share channel bandwidth.

5. Conclusion

We have designed an integrated-mode multi-channel architecture, called H-WLAN, for data dissemination in 802.11 WLAN info-stations. The main contributions of the proposed H-WLAN are as follows. First, we have shown that, by integrating the ad-hoc mode with the infrastructure mode, APs and ad-hoc nodes can exploit more channel resources to cooperatively provide a data dissemination service. Second, popularity-based channel allocation can efficiently allocate channel resources to object requests based on the estimated popularity of objects. Thus, the system can maximize multi-channel utilization based on the popularity of disseminated objects, and provide higher throughput. Third, mobile nodes use an adaptive source selection scheme to determine the best data source in a distributed manner, which can significantly reduce the workload on the info-station. In summary, our proposed schemes, *popularity-based channel allocation* and *adaptive source selection*, help balance the traffic load across 802.11 non-overlapping channels and available source candidates. Hence, users can get efficient and consistent service in the integrated-mode and multi-channel H-WLAN. In addition, since H-WLAN is built on the application level, it can be easily deployed and compatible with legacy IEEE 802.11 standard. Although the H-WLAN in this paper is designed for data dissemination applications, we believe that other applications could benefit by using the proposed H-WLAN framework.

Acknowledgements

This work was partially supported by the National Science Council and the Ministry of Education of ROC under the contract No. NSC95-2221-E-002-103-MY2, NSC95-2622E-002-018, and NSC96-2622-E-002-002.

References

- [1] A. Balachandran, G.M. Voelker, P. Bahl, Wireless hotspots: current challenges and future directions, *Mob. Netw. Appl.* 10 (3) (2005) 265–274.
- [2] M. Rodrig, C. Reis, R. Mahajan, D. Wetherall, J. Zahorjan, Measurement-based characterization of 802.11 in a hotspot setting, in: *E-WIND'05: Proceeding of the 2005 ACM SIGCOMM Workshop on Experimental Approaches to Wireless Network Design and Analysis*, 2005, pp. 5–10.
- [3] D. Ho, S. Valaee, Information raining and optimal link-layer design for mobile hotspots, *IEEE Trans Mobile Comput* 4 (3) (2005) 271–284.
- [4] D. Ho, S. Valaee, Mobile hotspots, In *Poster Session of INFOCOM 2005*, Miami, Florida, USA.
- [5] R.H. Frenkiel, T. Imielinski, Infostations: The joy of ‘many-time, many-where’ communications, WINLAB Technical Report (WINLAB-TR-119), Rutgers University, 1996.
- [6] R. Frenkiel, B.R. Badrinath, J. Borres, R. Yates, The infostations challenge: balancing cost and ubiquity in delivering wireless data, *IEEE Personal Commun.* 7 (2) (2000) 66–71.
- [7] T. Small, Z.J. Haas, The shared wireless infostation model: a new ad hoc networking paradigm (or where there is a whale, there is a way), in: *ACM MobiHoc'03*, 2003, pp. 233–244.
- [8] C.-J. Lin, C.-F. Chou, Hybrid wlan for data dissemination applications, in: *IEEE Computer Society's 12th Annual International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunications Systems (MASCOTS'04)*, 2004, pp. 113–120.
- [9] L. Yin, G. Cao, Supporting cooperative caching in ad hoc networks, *IEEE Trans. Mobile Comput.* 5 (1) (2006) 77–89.
- [10] P.U. WICAT, Information Project, <<http://wicat.poly.edu/>>.
- [11] T. Ye, H.-A. Jacobsen, R. Katz, Mobile awareness in a wide area wireless network of info-stations, in: *ACM MobiCom'98*, 1998, pp. 109–120.
- [12] Y. Sun, E.M. Belding-Royer, Application-oriented routing in hybrid wireless networks, in: *IEEE International Conference on Communications*, 2003. *ICC'03*, 2003, pp. 502–506.
- [13] Y. Bejerano, Efficient integration of multi-hop wireless and wired networks with qos constraints, in: *ACM MobiCom'02*, 2002, pp. 215–226.
- [14] W.H. Yuen, R.D. Yates, S.-C. Mau, Exploiting data diversity and multiuser diversity in noncooperative mobile infostation networks, in: *IEEE Infocom 2003*, vol. 3, 2003, pp. 2218–2228.
- [15] S.B. Kodeswaran, O.V. Ratsimor, A. Joshi, T. Finin, Y. Yesha, Using Peer-to-Peer data routing for infrastructure-based wireless networks, in: *First IEEE International Conference on Pervasive Computing and Communications (PerCom 2003)*, 2003, pp. 305–312.
- [16] J.-C. Chen, S.-H. Chan, J.-Y. He, S.-C. Liew, Mixed-mode wlan: The integration of ad hoc mode with wireless lan infrastructure, in: *IEEE Globecom'03*, vol. 1, San Francisco, CA, 2003.
- [17] R. Draves, J. Padhye, B. Zill, Routing in Multi-radio, Multi-hop Wireless Mesh Networks, in: *ACM MobiCom'04*, 2004, pp. 114–128.
- [18] A. Raniwala, T. Chiueh, Architecture and algorithms for an IEEE 802.11-based multi-channel wireless mesh network, in: *IEEE Infocom*, vol. 3, 2005, pp. 2223–2234.
- [19] M. Alicherry, R. Bhatia, L. Li, Joint channel assignment and routing for throughput optimization in multi-radio wireless mesh networks, in: *ACM MobiCom'05*, 2005, pp. 58–72.

- [20] J. Tang, G. Xue, W. Zhang, Interference-aware topology control and qos routing in multi-channel wireless mesh networks, in: ACM MobiHoc'05, 2005.
- [21] P. Kyasanur, S. Jungmin, C. Chereddi, N.H. Vaidya, Multichannel mesh networks: challenges and protocols, *IEEE Wireless Commun.* 13 (2) (2006) 30–36.
- [22] J. So, N.H. Vaidya, Multi-channel mac for ad hoc networks: handling multi-channel hidden terminals using a single transceiver, in: ACM MobiHoc'04, 2004, pp. 222–233.
- [23] A. Nasipuri, S. Das, Multichannel csma with signal power-based channel selection for multihop wireless networks, in: *IEEE Fall Vehicular Technology Conference*, vol. 1, Sep. 2000, pp. 211–218.
- [24] N. Jain, S. Das, A. Nasipuri, A multichannel csma mac protocol with receiver-based channel selection for multihop wireless networks, in: *Tenth International Conference on Computer Communications and Networks (ICCCN)*, 2001.
- [25] S. Kurkowski, T. Camp, M. Colagrosso, MANET simulation studies: the incredibles, *Mobile Comput. Commun. Rev.* 9 (2005) 50–61.
- [26] K. Pawlikowski, H. Jeong, J. Lee, On credibility of simulation studies of telecommunication networks, *Commun. Mag. IEEE* 40 (1) (2002) 132–139.
- [27] The Network Simulator – ns-2. URL <<http://www.isi.edu/nsnam/ns/>>.
- [28] J. Broch, D. Maltz, D. Johnson, Y. Hu, J. Jetcheva, A performance comparison of multi-hop wireless ad hoc network routing protocols, in: *Proceedings of the Fourth Annual ACM/IEEE International Conference on Mobile computing and networking*, 1998, pp. 85–97.
- [29] J. Nielsen, Zipf curves and website popularity (1997). URL <<http://www.useit.com/alertbox/zipf.html>>.



Kate Ching-Ju Lin received the B.S. degree in Department of Computer Science, National Tsing-Hua University in 2003. She is currently a Ph.D. student in Graduate Institute of Networking and Multimedia at National Taiwan University and also a member of Network Group in Communication and Multimedia Laboratory. Her research interests include wireless networks, multimedia communications, and peer-to-peer networks.



Cheng-Fu Chou received the M.S. and Ph.D. degree from University of Maryland, College Park, in 1999 and 2002, respectively. In 2002, he joined the Department Computer Science and Information Engineering, National Taiwan University. His current research interests are in wide-area network applications, distributed multimedia systems, heterogeneous wireless communication systems, and wireless sensor network.