

Station based fast handoff solution for IEEE 802.11 b/g wireless LANs

Ni Weiguo Dong Yongqiang Xia Qin

(Key Laboratory of Computer Network and Information Integration of Ministry of Education, Southeast University, Nanjing 210096, China)

Abstract: Real-time applications are sensitive to conditions such as transmission delay and jittering. To cut down on the influence generated by the WLAN handoff process, three parts of WLAN (wireless local area networks) handoff: handoff triggering, access point selection and the fast handoff algorithm are investigated. A fast handoff solution totally based on the station is proposed and it is composed of three parts: a handoff triggering mechanism based on dynamic threshold adjustment; an AP selection criterion based on probe delay; a fast handoff algorithm with differentiated channel selection and a dynamic cache. The station based solution is independent with AP's collaboration and avoids any changes in the IEEE 802.11 protocol. It is robust and has very good extensibility. Through tests and evaluation in a hotspot WLAN, the solution effectively reduces handoff latency and user experience of real-time applications is enhanced.

Key words: IEEE 802.11; WLAN; fast handoff

IEEE 802.11^[1] wireless local area networks (WLANs) have been broadly deployed and applied. Meanwhile, typical applications such as VoWLAN come out. WLAN^[2] is composed of APs (access points) and stations (wireless terminals). Interconnected with DS (distribution system), APs are the relay points through which stations make communication. To access the network, a station has to associate with an AP. An AP and its affiliated stations together form a BSS (basic service set), whose size is determined by the AP's signal strength coverage. Unlike traditional wired LANs, WLANs support mobility: you can move within one BSS or from one BSS to another. The interior mobility of a BSS does not trigger handoff, so we focus on mobility between APs. When a station moves out of its current AP's coverage and enters into another AP's coverage, a handoff will occur. To avoid losing network connectivity, a station has to perform handoff (BSS handoff): discover and select a new AP and authenticate itself and associate with it. Handoff brings in latency and has severe effects on real-time applications, so we have to optimize the handoff process and reduce the negative effects of handoff latency on real-time applications.

BSS handoff occurs on the link layer and it can be called MAC layer handoff or second layer handoff (L2 handoff). The process of BSS handoff is (see Fig. 1): when the network conditions fail to meet the communication requirements, a handoff is triggered; a station probes (scans) and selects a new AP from the results to perform authentication

and association; after a succession of associations, the station continues its communication through the new AP. This paper summarizes handoff into three parts: handoff triggering, AP selection and the process of handoff execution.

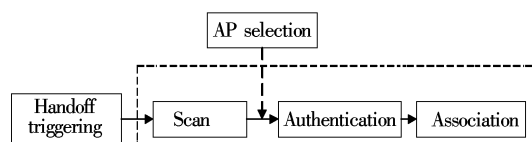


Fig. 1 BSS handoff process

According to IEEE 802.11 protocol, the WLAN handoff execution process comprises three phases:

1) Discovery phase

It is also named the probe phase (scan) where a station switches its working channel and probes on each channel to discover available APs in the vicinity. There are two ways of probing: passive and active. The former receives Beacon frames periodically broadcasted by APs, while the latter broadcasts probe request frames and receives probe response frames from APs.

2) Authentication phase

The authentication process between the station and AP. Several regulations are proposed for authentication including open, shared key, WPA and so on. The process is done with the exchange of authentication requests and authentication response frames or following the 802.1X protocol by a third party RADIUS server.

3) Association phase

The station completes association with AP via an exchange reassociation request frame and a reassociation response frame.

Each phase corresponds to latency, which is a discovery latency, an authentication latency and an association latency, respectively. The more complex the authentication process is, the longer the latency will be produced when handoff happens. For SIP-based applications like VoIP, the L2 authentication can be simple because these applications also have built-in application layer authentication. Present research on BSS handoff commonly holds a hypothesis that the L2 authentication method is either an open mode or a shared key. This paper reserves this opinion.

1 Fast Handoff

Handoff latency disrupts applications which are sensitive to delay and jittering, such as real-time audio applications. It is also the main problem still facing the massive popularization and deployment of real-time applications. BSS handoff has to be fast, smooth and have little influence on upper layer applications. Real-time applications have strict requirements for QoS parameters such as wireless link conditions,

Received 2007-10-26.

Biographies: Ni Weiguo (1982—), male, graduate; Dong Yongqiang (1973—), male, doctor, dongyq@seu.edu.cn.

Foundation item: The National Natural Science Foundation of China (No. 90604003, 60603067).

Citation: Ni Weiguo, Dong Yongqiang, Xia Qin. Station based fast handoff solution for IEEE 802.11 b/g wireless LANs[J]. Journal of Southeast University (English Edition), 2008, 24(2): 149 – 154.

so besides the execution process of handoff we have to pay attention to two other areas: handoff triggering and AP selection. The ultimate goal is to devise better handoff algorithms and optimize the handoff process.

1.1 Handoff triggering

The station is the decision maker of a handoff. When station observes that some predefined rules are violated, a handoff is triggered. Handoff triggering has a connection with handoff performance and it is a tradeoff between sensitivity and accuracy. There are two kinds of triggering mechanisms:

1) The station continually measures the signal strength value of RSSI (received signal strength indication) when communicating with AP; when RSSI is below a defined threshold, a handoff is triggered.

2) The decline in link quality will cause the retransmission of data frames or the failure to receive some frames, so handoff can be triggered if the frame retransmission number or the frame losing number exceeds a threshold. For example, as soon as the number of the consecutively losing beacon frames of the current AP is greater than a constant value, a handoff needs to be launched.

For real-time applications, the first way is too sensitive and may cause frequent station handoffs between APs. The second way is suitable for data applications but is not suitable for real-time applications. A new handoff triggering method still needs to be devised for real-time applications.

1.2 AP selection

Playing the role of data transmission for stations, AP performance directly touches user experiences. Station discovers APs by probing and selects one of them to associate with and gain access to the network. Currently the choice is made by an AP's RSSI and the "best" AP is the AP with the strongest signal strength. This simple method can lead too many stations choosing the same AP and the probability of medium collision and startup of a collision avoidance algorithm increases because stations together with their associated APs access the wireless medium under the rules of CS-MA/CA. A congested cell severely degrades user experience and as a whole, the network load is unevenly distributed. Signal strength is an aspect of link condition but it is not the entire delegation. The following is the related work aimed at improving AP selection:

- Making an extension of the Beacon frame or the probe response frame^[3-4] by adding metrics reflecting an AP's cell condition, such as the associating station number and the current network traffic on this AP. After obtaining these metrics, station can infer the AP ability by using some mathematical models and then select the "best" one of them. This is the most frequently recommended method.

- Cell breathing^[5]. When an AP finds itself overloaded, it just reduces the transmission power and forces some stations to handoff from it. This method just forces some stations to select other APs, and the adjustment of transmission power is usually difficult to do.

1.3 Fast handoff algorithm

Among the triple parts of the BSS handoff execution

process, the discovery phase contributes the greatest latency (about 90% of the total handoff latency^[6]). Many works have been done to improve the discovery phase. In order to decrease the handoff latency even more, a cache is introduced. A cache is an information store of APs in the vicinity; when handoff occurs, a station can directly select an AP from the cache and try to authenticate itself and associate with it. If the station fails to associate with all the APs in the cache, a normal discovery phase kicks in. Related works are:

Passive scanning (SyncScan^[7]): The station calculates the Beacon frame arrival time of every AP and switches to that AP's working channel at that time to receive Beacon frames for updating of AP information.

Active scanning: neighbor graph^[8]. A neighbor graph is built up using previous handoff experience and it is the topology of the APs that the station may handoff to. An entire neighbor graph can be stored on a centralized server^[9]. The station sends request to the server to obtain stored neighbor graph. The other is selective channel discovery^[10-11], probing channels selectively.

The above methods have good performance when the network topology is relatively simple and there is less station mobility. But the problem is that the cache contents are static and the maintenance only updates attributes of the AP in it. Under a mobile environment, it is hard to guarantee that APs in the cache are still available. When all the APs in the cache fail to authenticate themselves or associate, the station has to restart a normal discovery phase, so we have to solve the problem of low cache availability and increase its efficiency.

The aforementioned fast handoff algorithms and AP selection methods often need the AP's collaboration and their design and implementation modify the 802.11 protocol more or less. This modification makes it impossible for the mutual operation and compatibility of wireless network devices provided by different vendors. Sometimes it is difficult to update the existing WLANs. All the work of this paper is station-centered without any change to the protocol, so the solution and algorithm have very good compatibility and extensibility.

2 Station Based Fast Handoff Solution

The work of this paper is done only on the station and it is easy to carry the algorithm system to different device drivers. To solve the fast handoff problem for real-time applications, we make innovations on three aspects: a handoff triggering mechanism based on dynamic threshold adjustment; an AP selection criteria based on probe delay; a fast handoff algorithm with differentiated channel selection and dynamic cache.

Fig. 2 is a framework of our solution. From the startup to running, the WNIC (wireless network interface card) has gone through five states: INIT (initialization), SCAN (probing), AUTH (authentication), ASSOC (association) and RUN (running). The circles in Fig. 2 show our designed function modules:

1) Handoff triggering module. This module is for handoff triggering. When the WNIC enters into running state, it checks whether a handoff is needed through the course of

communication with the AP.

2) Scan module. It performs the discovery work by active probing and building up the backup AP cache using the probing results.

3) AP selection module. This module selects the “best” AP out of the cache with principles defined in it for the next process of authentication and association.

4) Cache maintenance module. The cache is a backup store of APs which is built up based on the results of probing. Available AP information is stored in the cache and in order to keep this information up-to-date, station does maintenance work periodically. After going into the running state, the station updates the cache recurrently.

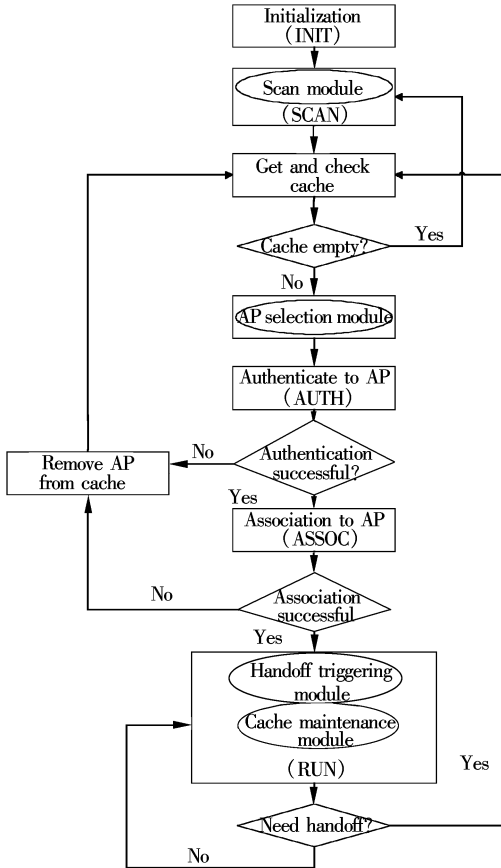


Fig. 2 Framework of station based fast handoff solution

2.1 Handoff triggering

Real-time applications have some requirements for link conditions, and steady link conditions that are important for them. For real-time applications, a station has to decide when to handoff accurately: not too late which can cause massive quality degradation of communication; not flapping between APs which can frequently cause handoffs. In consideration of sensitivity, we still use the handoff triggering mechanism according to the RSSI threshold but we make improvements in two ways:

1) Smooth processing of RSSI

To overcome the randomness and uncertainty of the RSSI value in instant time, we need to smooth RSSI values. The RSSI value at time $t + 1$ is computed by the RSSI value at time t , an RSSI sample value at time $t + 1$ and a factorial ∂ ($0 < \partial < 1$).

$$RSSI_{t+1} = RSSI_t \times \partial + Sample_{t+1} \times (1 - \partial)$$

2) Dynamic handoff triggering threshold

For the elimination of frequent handoff problems, we introduce dynamic handoff triggering thresholds. Three thresholds are defined:

- ① Normal_threshold: the floor level of the RSSI value when normally communicating with the AP;
- ② Prime_handoff_threshold: the prime handoff threshold;
- ③ Second_handoff_threshold: the second handoff threshold. Among them, ① > ② > ③.

At the initial time, the handoff triggering threshold is ②; when handoff occurs and after the station successfully hands off to another AP, adjust the threshold to ③; when in running state and RSSI value of associated AP is not below ①, adjust the threshold back to ②.

Smoothing eliminates the uncertainty of the RSSI value and the dynamic adjustment of the triggering threshold prevents frequent handoffs from occurring; both help the station make accurate handoff decisions.

2.2 Access point selection

Selection of AP only based on the RSSI may lead many stations to focus on some specific APs. The congested link condition, in turn, will degrade network performance and distribute network load unevenly. In two circumstances the station need to select AP: when station boots up and when a handoff is going on. We conclude the two as one because they both need to select AP from cache originally built up or updated by maintenance. Both built up and maintenance of cache are fulfilled through active probing, so we can use the delay of the probing process.

The exchanges of probe request and probe response frames like other frames follow the CSMA/CA rule to obtain the medium. Congested wireless link increases the transmission delay of management frames like what happens to the data frames. We define the concept of probe delay: time between the reception of probe response and the transmission of probe request.

$$T_i = T_{\text{recv}_i} - T_{\text{send}}$$

where T_{recv_i} is the time when receiving the i -th probe response from AP; T_{send} is the time when broadcasting the probe requests.

When an AP is congested and overloaded, the sending delay of probe responses is increased. The IEEE 802.11 protocol defines the time interval a station has to wait on each channel when probing for AP discovery. There are two intervals: minChanTime and maxChanTime. If the time of minChanTime has elapsed since the time sending probe request and no probe responses have been received, the station switches to the next channel to go on probing; otherwise, the station continues to stay on that channel to receive more probe responses until maxChanTime expires. Usually more than one response can be received from each AP and we take the first probe delay as the most important, so we calculate the probe delay as follows:

$$T_{\text{probe_delay}} = T_1 + \frac{\sum_{i=2}^N (T_i - T_{i-1})}{N - 1}$$

According to the formula, the $T_{\text{probe_delay}}$ of AP is the sum of two parts: one is the first probe delay T_1 , the other is the arithmetic average of variance between two adjacent probe delays. When the station makes a selection of APs, it compares the probe delay of APs with priorities. Meanwhile, in order to obtain better communication quality, the station also considers the RSSI of the AP. Here is the AP select algorithm:

Algorithm 1 AP selection

```

Broadcast probe requests on each channel;
Receive probe responses and derive  $T_{\text{probe\_delay}}$  for each AP;
...
AP(target) = AP(1);
for AP(i) in AP(2) ... AP(M)
{
  if( AP(target).  $T_{\text{probe\_delay}}$  > AP(i).  $T_{\text{probe\_delay}}$ 
    && |AP(target). RSSI - AP(i). RSSI| ≤ Δ)
    AP(target) = AP(i)
}

```

Notes: AP(target) is the object AP we want to choose; Δ is a constant indicating the upper limit of RSSI variance.

2.3 Fast handoff algorithm

Both handoff triggering and AP selection are parts of the handoff process. It is the fast handoff algorithm that improves the handoff performance and reduces the handoff latency. Presently two areas still need to be worked on: improving the efficiency of channel probing and enhancing the availability of APs in the cache. For them, two new solutions are proposed: probing by differentiated channel selection and dynamic cache assisted AP information maintenance. We cover the two in detail respectively below.

2.3.1 Probing by differentiated channel selection

There are only three channels without overlapping with each other of all the eleven channels given by the IEEE 802.11 b/g protocol; channels 1, 6, and 11 (in different countries the number may be different; we take the regulation of FSM in America as an example). Overlapping channels interfere with each other and influence mutual communication. When we deploy a hotspot WLAN, we select channels only among the three and through network planning to prevent adjacent APs working on the same channel. We divide the channels into halves:

Static channel set(S_chan_set): channels 1, 6 and 11;

Dynamic channel set(D_chan_set): all the channels but 1, 6 and 11.

We process the channels in the two sets differently: channels in S_chan_set are immune to change but the channels in D_chan_set are changing with the probing results.

The time the station stays on one channel when probing is determined by two time intervals: minChanTime and maxChanTime. When minChanTime expires and the current channel is in D_chan_set , we remove the channel from set. By preferentially probing the non-overlapping channels and

a dynamic adjustment of the channel set, the station can accurately discover APs and jump over channels that have no AP on them.

The worst case is that if the station has probed all the channels in the two sets and still no APs have been found, then D_chan_set will be reset and restart the probing.

2.3.2 Dynamic cache assisted AP information maintenance

The introduction of a cache is an effective way to cut down on handoff latency. When a handoff occurring, the station first uses APs in the cache and no more probing work is needed. If the station successfully associates with an AP in the cache, handoff latency decreases dramatically. But if all the APs in the cache fail to associate, the station has to restart the normal probing phase and thus increases the handoff latency in the opposite directions. Cache maintenance, occurring periodically, temporarily stops transmission and reception of data frames and cost is included. So the availability of APs in the cache is very important.

The problem is that all the cache contents are static so that we only update attributes of APs in the cache and do not add any new APs to the cache or delete any APs from the cache. The contents are changeless unless all the APs fail. But when the station moves, the APs in its vicinity change; that is, the station leaves some old AP coverage and enters some new AP coverage. This is common because real-time applications require better signal coverage of the WLAN and as long as the channels are different, the coverage of two APs can become deeply overlapped. Since some APs are not available anymore, we should replace these APs in the cache with other newly encountered APs.

In this paper, we propose a dynamic cache. The maintenance of the cache should face the channels that the station probes channels in S_chan_set and D_chan_set (Channels that have no APs being worked on are already removed). If some new AP has been found and is better than an AP in the cache working on the same channel, the old AP will be replaced by the new one. If the AP is already in the cache, the AP's information is updated. When the station is moving, the APs surrounding the station are also changing; the APs in the cache are also changing synchronously. The station temporarily leaves its working channel when doing maintenance work and to avoid frame loss, the station informs the current AP that it will enter into "power saving" mode before switching to other channels. After fulfillment of cache maintenance, the station retrieves the data buffered at the AP by sending PS poll frames.

Algorithm 2 Fast handoff algorithm

```

while( true) {
  sample and smooth RSSI;
  if( RSSI < Handoff_Threshold
    && ieee80211_state == RUN) {
    while( get and check cache) {
      if( cache is not empty) {
        choose AP(i) with AP selection algorithm;
        try authenticate and associate with AP(i);
        if( successful with AP(i)) {
          Handoff_Threshold
            = second_Handoff_Threshold;

```

```

        break;
    }
    else{
        delete AP(i) info;
    }
}
else {
    reset D_chan_set;
    probe chan consecutively;
    update the cache;
}
}
}
else if(RSSI >= normal_Threshold
    && ieee80211_state == RUN)
    Handoff_Threshold =
        prime_Handoff_Threshold;
}

```

3 Solution Testing and Analyses

We implement all the three parts of our fast handoff solution on wireless terminals, and experiments are done in a typical hotspot WLAN for analysis and evaluation.

3.1 Test bed and related configuration

The hotspot is an IEEE 802.11 b/g WLAN with 12 APs distributed among a building floor and all the APs are working on channel 1, 6 or 11. In our experiment, we use 6 of them and the topology of the APs is given in Fig. 3. The APs are positioned on the ceiling and have one or two antennas reaching out to their coverage area.

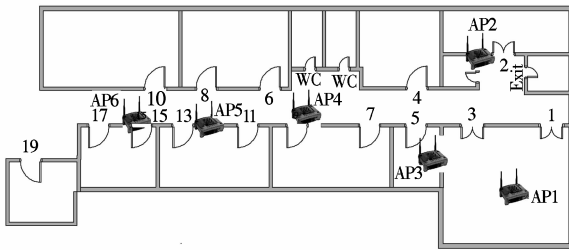


Fig. 3 Hotspot WLAN for experiment

The mobile node is a HP Compaq nc4200 laptop with Linux (Fedora kernel version 2.6.15) as operating system and equipped with TP-Link WN510G cardbus WNIC. We choose Madwifi as our developing platform which supports WNIC built in with Atheros chipsets. The version of the open source driver is madwifi-ng.

Using Madwifi, we implement our fast handoff solution on the driver of WNIC. The three handoff triggering thresholds are 20, 8 and 5, respectively. We use the kernel timer to perform cache maintenance work and the timer expiration interval is 30 s. The timer is activated when the station is in running state and it is stopped when a handoff is triggered. Two important values for active probing are: minChanTime is 7 ms, and maxChanTime is 11 ms. To have records related to handoffs, we write time of handoff triggering, handoff latency and time when starting to update the cache into the

system log. The time is a relative value with the time when the WNIC is activated as a base point.

3.2 Experiments and analyses of performance

To simulate the data flow of VoIP applications, we use the network diagnosis tool “ping” to send ICMP requests to a host on an Ethernet every 20 ms and receive ICMP responses from the corresponding host. We record the ICMP RTT (round trip time) at the laptop and capture the ICMP request packets using Ethereal at the host side which are used to calculate the inter-arrival time of ICMP requests. To obtain reliable experimental result and eliminate influences such as MAC address learning of switchers, we omit records before the 100th ICMP packets. To obtain true experimental data, the experiments are often practiced on the weekend or in the evening with very few interfering origins and we try to keep moving along the same way, at the same speed during each experiment.

The results are shown in Fig. 4 and Fig. 5. In Fig. 4 we can see that handoff occurs when a wireless link condition is deteriorated which is reflected by the increase in RTT delay. The latency produced by cache maintenance is less than 50 ms with an average of about 30 ms (usually probing 3 channels). The average handoff latency is about 10 ms because of the improvement in cache availability. Fig. 5 reveals the inter-arrival time of ICMP packets at the corresponding host. Although handoff occurs 3 times, there are very few packets jittering and they account for less than 1% of all the packets. We can also observe that some packet inter-arrival times are 0; the reason is that when the station updates a cache, it not only informs the AP to buffer downlink data but also creates a local buffer for uplink data on itself. When the maintenance of the cache is completed, it immediately clears up the local buffer and sends pending data without considering the time when these data are passed down from the upper layers, so some ICMP packets arrive at the corresponding side at the same time.

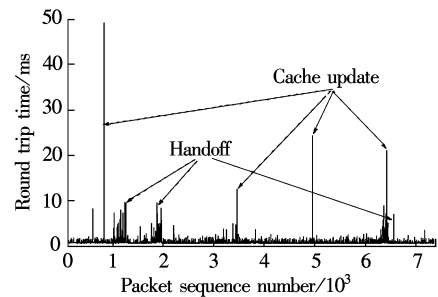


Fig. 4 RTT time of ICMP packets

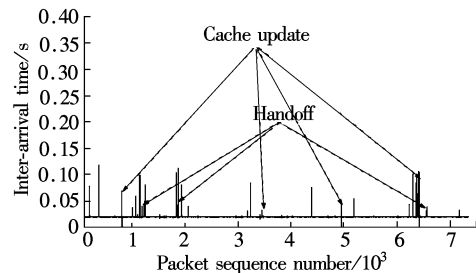


Fig. 5 Inter-arrival time of ICMP packets

4 Conclusion

Real-time applications on WLAN are sensitive to handoff latency and one of the preconditions of large scale deployment of real-time applications is reducing handoff latency and optimizing the handoff process. First we analyze the handoff procedure and investigate three aspects of the handoff: handoff triggering, AP selection and the fast handoff algorithm. To keep protocols consistent and to get better extensibility of our solution, we design and implement all our proposals on station only. The station-centered fast handoff solution is composed of three parts: a handoff triggering mechanism based on a dynamic threshold adjustment; an AP selection criteria based on probe delay; a fast handoff algorithm with differentiated channel selection and a dynamic cache. Through experiments on the hotspot WLAN, our solution has better handoff performance and strong extensibility. It is very easy to deploy our solution onto the existing Hotspot WLANs.

References

- [1] IEEE Standard 802.11. Wireless LAN medium access control (MAC) and physical layer (PHY) specifications [S]. LAN/MAN Standards Committee of the IEEE Computer Society, 1999.
- [2] Gast Matthew S. 802.11 wireless networks—the definitive guide [M]. Beijing: Tsinghua University Press, 2002: 77 – 93.
- [3] Sundaresan Karthikeyan, Papagiannaki Konstantina. The need for cross-layer information in access point selection algorithms [C]//*Proc of the 6th ACM SIGCOMM on Internet Measurement*. New York, NY, USA: ACM, 2006: 257 – 262.
- [4] Abusubaih Murad, Gross James, Wiethoelter Sven, et al. On access point selection in IEEE 802.11 wireless local area networks [C]//*Proc of 31st IEEE Local Computer Networks Conference*. Tampa, FL, USA, 2006: 879 – 886.
- [5] Brickley O, Rea S, Pesch D. Load balancing for QoS optimization in wireless LANs utilising advanced cell breathing techniques [C]//*Proc of IEEE 61st Vehicular Technology Conference*. Stockholm, Sweden, 2005, 3: 2105 – 2109.
- [6] Shin M, Mishra A, Arbaugh W. An empirical analysis of the IEEE 802.11 MAC layer handoff process [J]. *ACM SIGCOMM Computer Communication Review*, 2003, 33(2): 93 – 102.
- [7] Ramani I, Savage S. SyncScan: practical fast handoff for 802.11 infrastructure networks [C]//*Proc of IEEE INFOCOM*. San Diego, La Jolla, CA, USA, 2005, 1: 675 – 684.
- [8] Shin M, Mishra A, Arbaugh W. Improving the latency of 802.11 hand-offs using neighbor graphs [C]//*Proceedings of the 2nd International Conference on Mobile Systems, Applications, and Services*. Boston, MA, USA, 2004: 70 – 83.
- [9] Park S, Kim H, Park C, et al. Selective channel scanning for fast handoff in wireless LAN using neighbor graph [C]//*Lecture Notes in Computer Science*. Springer, 2004: 194 – 203.
- [10] Liao Yong, Gao Lixin. Practical schemes for smooth MAC layer handoff in 802.11 wireless networks [C]//*Proceedings of the 2006 International Symposium on World of Wireless, Mobile and Multimedia Networks*. Washington, DC, USA: IEEE Computer Society, 2006: 181 – 190.
- [11] Shin S, Forte A G, Rawat A S, et al. Reducing MAC layer handoff latency in IEEE 802.11 wireless LANs [C]//*Proc of the Second International Workshop on Mobility Management & Wireless Access Protocols*. Philadelphia, PA, USA, 2004: 19 – 26.

基于站的 IEEE 802.11b/g 无线局域网快速切换方案

倪维国 董永强 夏 勤

(东南大学计算机网络和信息集成教育部重点实验室, 南京 210096)

摘要: 由于实时应用对传输时延和抖动相对敏感, 为减小 WLAN 切换过程对实时应用的影响, 研究了 WLAN 切换过程的 3 个方面: 触发切换, AP 选择和快速切换算法. 充分考虑到方案的可部署性和可扩展性, 提出了一个以终端站为中心的快速切换方案. 设计实现了基于动态阈值的切换触发机制、基于探寻时延的 AP 选择机制和信道区分选择结合动态缓存的快速切换算法. 该方案完全在站上实现, 不需要 AP 的配合或者对协议进行修改, 因而具有良好的鲁棒性和可部署性. 通过在实际的 hotspot WLAN 环境下测试和分析, 该快速切换方案有效地减小了切换时延, 同时使用户获得的实时应用网络性能有所增强.

关键词: IEEE 802.11; 无线局域网; 快速切换

中图分类号: TP391