A Dynamic Load Balancing Scheme for VoIP over WLANs

Chih-Chien Hsu and Shiao-Li Tsao
Dept. of Computer Science, National Chiao Tung University,
Hsinchu, Taiwan, R.O.C.
{hsucc, sltsao}@csie.nctu.edu.tw

ABSTRACT

Coverage areas of WLAN access points (APs) are usually overlapped so a WLAN station (STA) might find several APs to attach in a WLAN hotspot. Generally speaking, a WLAN STA associates with an AP with the maximal signal strength, and requests the bandwidth of the AP for establishing network connections. However, this kind of STA-centric association and bandwidth request policy may introduce unbalance loads of APs, and the utilizations of APs in a WLAN hotspot cannot be maximized. This unbalance load problem for a WLAN hotspot becomes one of the critical issues for the commercial deployment of WLAN services such as voice over IP (VoIP) over WLAN (VoWLAN) which needs to guarantee the Quality of Service (QoS) and also to satisfy the maximal number of subscribers. In this work, a novel dynamic load balancing scheme that works together with a network-assist association policy is proposed for a VoWLAN system. The network-assist association policy first advices an STA to request a VoWLAN session through an AP with the minimal load. In case the APs that the STA can attach are all overloaded, the proposed load balancing scheme further rearranges the serving VoWLAN STAs between APs in order to spare enough resources to service that new request. Simulation results demonstrate that the utilizations of APs in a WLAN hotspot are improved by applying the proposed scheme.

Keywords - voice over WLAN (VoWLAN), radio resource management, load balancing, voice over IP (VoIP), WLAN

1: INTRODUCTIONS

The technology development and network deployment of WLANs both grow dramatically in recent several years and now WLAN becomes one of the most popular access technologies for mobile Internet services. Among all mobile Internet services and applications, voice over IP (VoIP) over WLAN (VoWLAN) attracts great interests and is regarded as a killer application for both public and enterprise WLANs [1]. However, VoWLAN applications generate a large amount of small voice packets which degrade the WLAN utilization due to the nature of WLAN medium access control (MAC) mechanism [7], and the service capacity of a WLAN AP for VoWLAN services, i.e. the number of concurrent VoIP sessions that a WLAN AP can support, is very limited [2]. Previous studies have been working on the improvement of the utilization of a single WLAN AP for VoWLAN services [3], and also proposed several radio resource management schemes for WLANs [4][5][6]. Bejerano et al further consider the load balancing issues between APs and present a fairness bandwidth allocation policy for STAs [10]. However, subscribers in a VoWLAN system may request different QoSs and bandwidths of APs, and bandwidth and QoS guarantee are more important than the fairness bandwidth allocation for STAs in a VoWLAN application. Their proposed mechanism cannot be directly applied to a VoWLAN system.

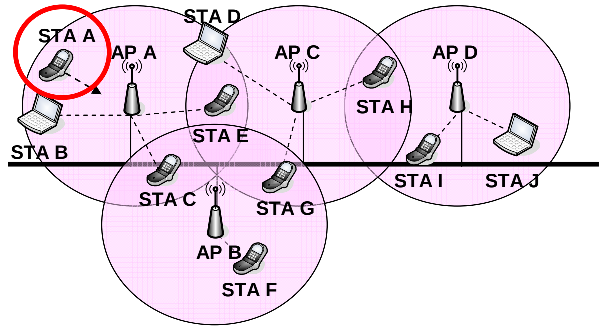
The coverage areas of WLAN APs are usually overlapped and a WLAN STA might find several reachable APs while a WLAN STA joins a WLAN hotspot network. Normally, a WLAN STA selects an AP with the maximal signal strength to associate and to establish connections. Then, the STA occupies certain resources such as bandwidth and buffer of the AP for the connections and services. However, this STA-centric network association and service request policy may introduce unbalance loads on APs, and the utilizations of APs in a hotspot cannot be maximized. This issue is especially important for these applications such as VoWLAN services that need to support a large number of sessions with quality of services (QoSs). To manage WLAN resources from a network point of view, the IEEE 802.11k that offers radio measurement and management facilities on APs and STAs is thus proposed [8]. However, mechanisms to balance the loads between APs for VoWLAN applications are not yet fully elaborated in previous studies and the specifications. In this work, a novel dynamic load balancing scheme that works together with a network-assist association mechanism is proposed for a VoWLAN system. The network-assist association mechanism advices an STA to associate with an AP with the maximal available resources. In case of the APs that the STA can attach are all overloaded, the proposed load balancing scheme is activated to adjust the loads between APs in order to accommodate that new VoWLAN request.

The rest of the paper is organized as follows. The concept and procedures of the proposed dynamic load balancing scheme for a VoWLAN system are presented in Section II. Simulation results are discussed in Section III, and finally Section IV concludes this work.

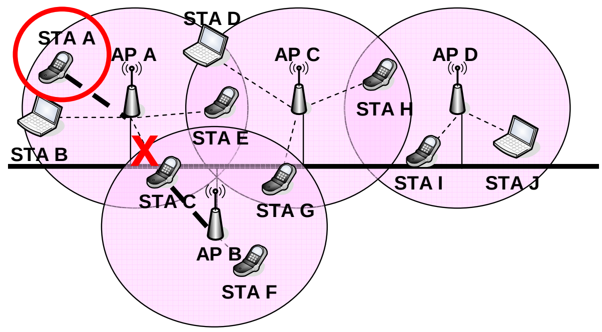
2: DYNAMIC LOAD BALANCING SCHEME

2.1: Dynamic load balancing example

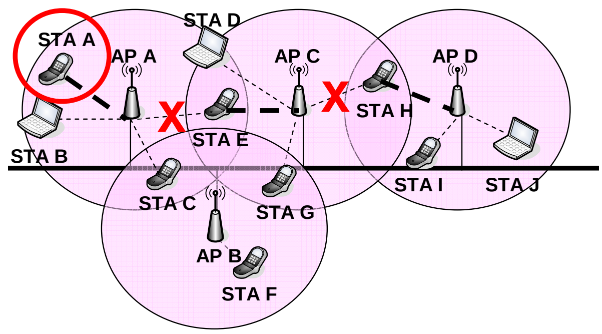
An STA, say STA A, is able to associate with a WLAN AP, say AP A, only. While AP A is overloaded, the STA A cannot obtain enough resources from AP A and the service request from STA A is thus rejected by AP A. Considering AP A is currently serving another VoWLAN STA, say STA B, and STA B can find and associate with another AP, say AP B, which is under-loaded. STA B can change its serving AP from AP A to AP B, and then the resources occupied by STA B on AP A can be released. Therefore, the resources on AP A now become available to serve the new STA A. The load adjustment procedure can be done between two APs, and it can be also applied to adjust loads between multiple APs. Figure 1 (a) shows an example where circles represent the coverage areas of APs and the adjoined APs are assumed to occupy different WLAN channels. In this example, each AP is assumed to support at most three VoIP sessions. STAs B, C, D, E, F, G, H, I and J associate with APs A, A, C, A, B, C, C, D and D respectively. While STA A that can only associate with AP A attaches to the network and requests VoWLAN services, AP A which is overloaded can not provide the service. A dynamic load balancing scheme is thus applied to the situation. The scheme changes the serving AP of STA C from AP A to AP B, and then the resources on AP A become available to be allocated to STA A. Therefore, the service request from STA A can be accepted by AP A. Figure 1 (b) illustrates the example shown in Figure 1 (a) after applying the proposed dynamic load balancing. Figure 1 (c) shows another example that the proposed dynamic load balancing can be performed as a chain. While STA A requests a VoIP session to AP A, and AP A is overloaded. STA H can change its AP from AP C which is also overloaded to AP D which is under-loaded. AP C has available resources to serve new requests, and then STA E can change its serving AP from AP A to AP C. Therefore, the resources on AP A become available to be allocated to STA A. In this study, STAs and APs are assumed to implement the mechanisms such as 802.11k so that STAs can be forced to change the current serving AP. Also, the migration procedure performs just like an inter AP handover and is assumed not to introduce too much packet delay and loss for VoIP sessions.



(a) Before load adjustment, STA A cannot obtain the resources



(b) After migrating STA C from AP A to AP B, STA A can obtain the resources



(c) After migrating STA H from AP C to AP D, STA E from AP A to AP C, STA A can obtain the resources.

Figure 1. A dynamic load balancing example

2.2: Model of relationships between APs and STAs

Before the dynamic load balancing scheme is described, the relationships between STAs and APs in a WLAN hotspot are first modeled. A WLAN hotspot totally contains N WLAN APs, and all APs in the hotspot are assumed identical. The current resource utilization of i^{th} AP, say A_i, is denoted as C_i which is a value between 0 and 1. If C_i=1, that means all resources on A_i are occupied by VoWLAN STAs and there is no resource available to serve any new STA. The j^{th} STA, denoted as S_j, associates with a WLAN AP, say A_i, at $R_{i,j}$ speed in Kbps. For example, the IEEE 802.11b offers 1Mbps, 2Mbps, 5 Mbps, and 11Mbps speeds for STAs, and the associated speed depends on the distance and channel quality between the AP and the STA. Assume an STA is requesting a VoWLAN session at r Kbps, an AP then allocates $r/R_{i,j}$ resources for a VoWLAN session if A_i admits S_j. The above equation uses a very simple model to evaluate the resource consumed by an STA associating with an AP at $R_{i,j}$ speed and requesting r bandwidth for its VoIP session. The resource consumption model can be more complicated and complete than this model. For example, the IEEE 802.11e suggests a resource consumption model while performing the call admission control [9]. Any resource consumption model and call admission control mechanism for WLANs can be employed on the proposed mechanism. Without loss of generality, the simple model is used to describe the basic idea behind the proposed load adjustment scheme. Besides the resource models, two relationships between APs and STAs are also defined. While an STA, say S_j, performs WLAN channel scan and finds an AP, say A_i, S_j then inserts A_i into the candidate list. Here, $n_{i,j}$ defines the coverage relationships between APs and STAs as:

$$n_{i,j} = \begin{cases} 1, & \text{if } A_i \text{ is in } S_j\text{' candidate list.} \\ 0, & \text{otherwise.} \end{cases}$$

After scan procedures, S_j decides to associate with A_i, and requests a VoWLAN service, $m_{i,j}$ defines the serving relationships between APs and STAs as:

$$m_{i,j} = \begin{cases} 1, & n_{i,j} = 1 \text{ and } A_i \text{ is serving } S_j. \\ 0, & \text{otherwise.} \end{cases}$$

For example, for STA E in Figure 1(a), $n_{A,E} = 1$, $n_{C,E} = 1$ and $m_{A,E} = 1$, $m_{C,E} = 0$.

2.3: Dynamic load adjustment

To achieve a better network utilization, a network-assist association policy assigns the AP with the minimal load to serve a new VoWLAN STA. That is, a new VoWLAN STA, say S_j, is asked to associate with A_i that is in S_j's candidate list, i.e. $n_{i,j}=1$, and A_i has the maximal available resources after serving S_j among all APs in S_j's candidate list, i.e. A_i with the minimal $C_i+r/R_{i,j}$. If a VoWLAN STA cannot find an AP with enough radio resources from its candidate list, the traditional approach rejects this VoWLAN STA. In this study, the second step procedure, i.e. dynamic load balancing procedure, is activated to adjust loads between APs to accommodate that new request. A direct graph represents the current loads and the relationships between APs and STAs. The directed graph, called resource-allocation graph ***G*** illustrates the resource-allocated status between APs and STAs. Vertices in graph ***V*** could be STAs or APs. An edge ***E*** denotes the relationship between vertices, and it only exists between one AP and one STA, but does not exist between two APs or two STAs. An edge from A_i to S_j denotes as (A_i, S_j) means AP A_i is serving STA S_j, called an assignment edge. That is $n_{i,j}=1$ and $m_{i,j}=1$. If there is an edge from S_j to A_i, denoted as (S_j, A_i), implies A_i is in S_j's candidate list but A_i is not serving S_j, called a claim edge. In other words, $n_{i,j}=1$ and $m_{i,j}=0$. Figure 3 shows an example of the resource-allocation graph for Figure 1 (a).

The relationship between APs and STAs can be easily obtained from the resource-allocation graph, and the graph will be also used in the dynamic load balancing scheme to determine the load adjustments between APs.

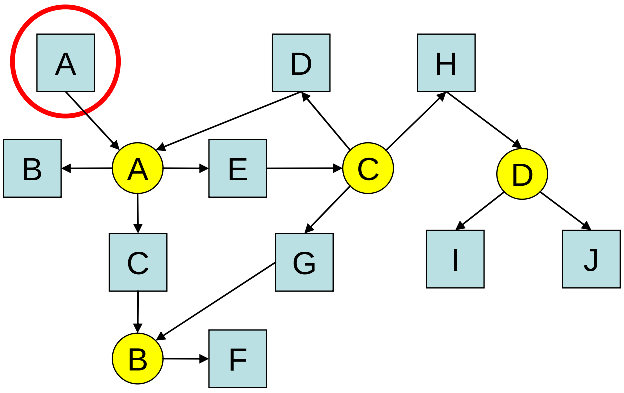


Figure 3. The resource-allocation graph of Fig. 1(a)

It can be seen from Figure 3 that STA A can be only served by AP A but unfortunately AP A is overloaded. The next step of load balancing procedure is to find a directed path without an STA traced twice in the resource-allocation graph from STA A to any other AP with available resources. A feasible path ***P*** in ***G*** for S_j is defined as $\{(S_j, A_i), (A_i, S_j'), (S_j', A_i'), \ldots, (S_j^{(n)}, A_i^{(n)})\}$. In this path, A_i to $A_i^{(n-1)}$ are overloaded and only $A_i^{(n)}$ is under-loaded and can serve $S_j^{(n)}$. A path represents a list of load adjustment operations. For example, S_j' can change its current serving AP from A_i to A_i', and then A_i has available resources to serve S_j. Before the migration of S_j', S_j'' can changes its serving AP from A_i' to A_i''. Since $A_i^{(n)}$ is under-loaded, STAs S_j' to $S_j^{(n)}$ can perform migrations for the old serving APs to the new serving APs. Therefore, the new STA, S_j, can be admitted and served by A_i. A VoWLAN STA might have multiple feasible paths. For example, the case shown in Figure 3 has at least two feasible paths, i.e. {(STA A, AP A), (AP A, STA C), (STA C, AP B)} and {(STA A, AP A), (AP A, STA E), (STA E, AP C), (AP C, STA H), (STA H, AP D)} If more than one path is found, the shortest path which implies the minimal migration overheads is selected. Once the path is decided, the direction of edges of the path should be reversed. That is, assignment edges become claim edges and claim edges become assignment edges. Figure 4 illustrates the resource-allocation graph of Fig. 1(c) after the load adjustment is performed. While STA A cannot get the resources from AP A, the dynamic load balancing mechanism is to find a load adjustment path ***P*** from STA A to AP D: {(STA A, AP A), (AP A, STA E), (STA E, AP C), (AP C, STA H), (STA H, AP D)}. Then, the directions of the edges in the path should be reversed. The algorithm of the proposed dynamic load balancing scheme is presented in Figure 5.

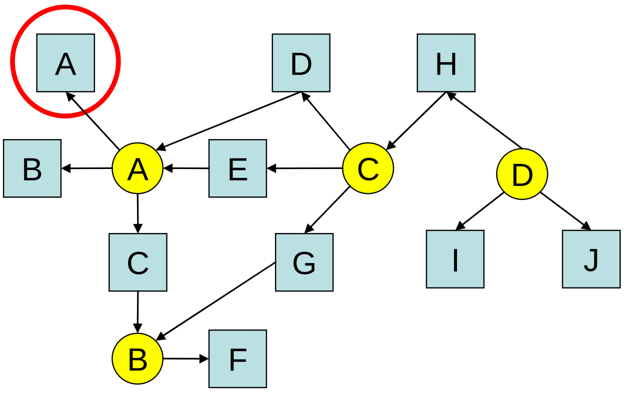


Figure 4. The resource-allocation graph of Fig. 3 after performing load adjustments

```
CP: set of feasible paths.
P: a feasible path.
SP: the shortest feasible path.
V: set of vertices that are visited.

Dynamic_Load_Balance (S_j)
{ if (Initial_Assignment(S_j) = false)
  { if ((CP=Find_Path(S_j, R_i,j)) = true)
     {SP=Find_Shortest_Path(CP);
     //return the shortest path.
     Reverse_Edge(SP);
     //reverse the directions of the
edges in a path.
     }
  }
}

Initial_Assignment(S_j)
{min_C=1;
for i where n_i,j=1
  {if C_i+r/R_i,j <= 1 and C_i+r/R_i,j < min_C
    min_C=C_i+r/R_i,j and min_A=i
  }
if min_C=1 return false
else
return min_A //return AP with minimal
load after serving the VoWLAN STA.
}

Find_Path(S_j,   R_i,j)//   depth   first
traversal
{
Add S_j to V;
for i where n_i,j=1 and m_i,j=0 AND A_i ∉ V
 {Add A_i to V;
  Add (S_j,A_i) to P;
//Search for STA S_k from the graph where
m_i,k=1
  if (C_i+r/R_i,j<=1)
     Add P to CP;//path is found
  else if( S_k ∉ V and C_i+r/R_i,j-r/R_i,k<=1){
     {Add (A_i,S_k) to P;
     Find_Path(S_k,R_i,k);
     Delete (A_i,S_k) from P;}
   Delete A_i from V;
   Delete (S_j,A_i) from P;}
Next
   Delete S_j from V;
}
```

Figure 5. The proposed dynamic load balancing scheme

3: SIMULATION RESULTS AND ANALYSES

Simulations are conducted to evaluate the utilizations of APs in a WLAN hotspot by applying the traditional STA-centric association mechanism and the proposed scheme. In the simulations, different numbers of APs, each covers a 30-meter-radius range, are randomly deployed over a fixed-size area which is 300 meters by 300 meters. To simplify the simulations, only one association speed, i.e. 11Mbps, is offered within an AP's coverage. Also, STAs are randomly distributed in the WLAN hotspot. Assume G.711 which requires total 160Kbps bandwidth is used for the voice codec in a VoWLAN session, and thus each AP can support up to eight VoWLAN sessions concurrently. The simulations also assume that an STA associates with an AP and only initiates one VoWLAN session through the AP. We aim to evaluate the network utilization, *U*, which is defined as the number of VoWLAN STAs served by all APs dividing by the total capacity of the APs in the hotspot. For example, a WLAN hotspot has 10 APs, and the total service capacity is 80 VoWLAN STAs, i.e. 80 VoWLAN sessions. If *U*=0.8, the system is currently handling 64 VoWLAN STAs. The next three simulations fix the number of APs in a hotspot to 10, 50 and 100, and investigate the utilizations of APs while different number VoWLAN STAs arrive to the WLAN hotspot.

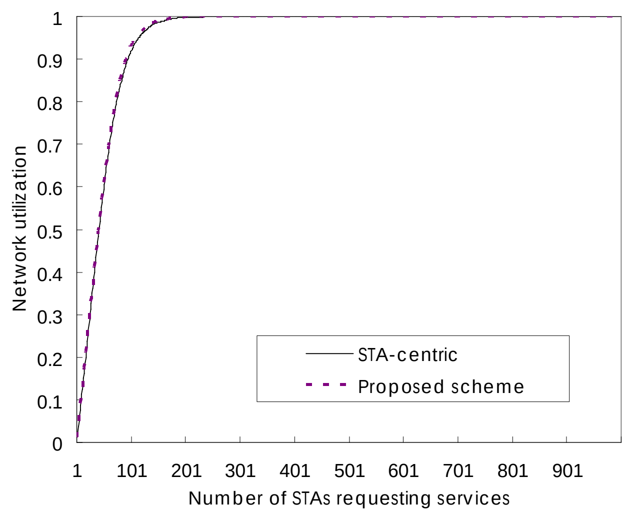


Figure 6. Network utilization for a WLAN hotspot with 10 APs

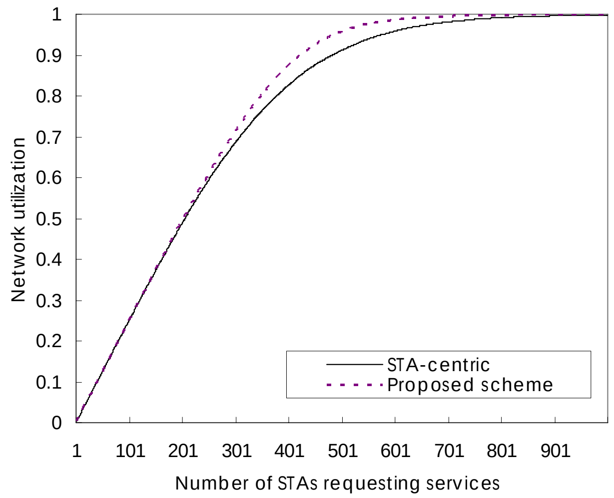


Figure 7. Network utilization for a WLAN hotspot with 50 APs

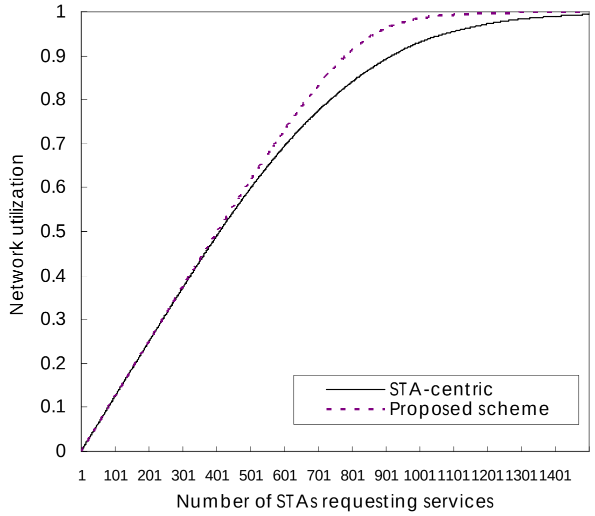


Figure 8. Network utilization for a WLAN hotspot with 100 APs

Figure 6 shows the utilizations of APs over the number of STAs requesting VoWLAN services. This simulation assumes the number of APs in a hotspot is 10, and in order words, a total 80 VoWLAN STAs can be served in the system at most. We can learn from Figure 6, while the number of STAs requesting the VoWLAN services is 100, and the utilization is still 92%. That implies only 74 VoWLAN STAs are admitted, and the other 26 VoWLAN STAs are rejected. Also, there are available resources on some specific APs to handle other 6 STAs. The reason why the system capacity of the hotspot is 80 STAs but only 74 STAs are admitted is because some VoWLAN STAs request the same APs which are already overloaded. Figure 6 reveals while the number of STAs requesting the services is over 250, the APs in a hotspot are fully utilized and 80 STAs can be served. The objectives of the proposed dynamic load balancing scheme are to minimize the service rejection rate and to maximize the utilizations of APs while the system suffers from a heavy load.

Figure 6 show that the proposed method performs almost the same as the traditional approach. This is because the proposed approach can improve the system utilization only when the coverage areas of APs have overlaps. Therefore, STAs situated in the overlapped areas can change their serving APs and the loads on APs can be adjusted to accommodate more STAs. In this simulation that only 10 APs are in the 300 meters by 300 meters area, the overlapped area is rare so that the dynamic load balancing seldom finds load adjustment paths. In Figure 7, the number of APs increases to 50. The results indicate that the proposed approach can improve the utilization by around 6%. In other words, the system by employing the proposed approach can admit 6% more VoWLAN STAs than the traditional approach under a heavy system loading such as 440 VoWLAN STAs request services. Obviously, for a hotspot area with 50 APs, the overlaps between APs increase, and more VoWLAN sessions can be admitted by applying our scheme. Figure 8 further increases APs to 100 which an STA can find two or three APs in this case, and the proposed scheme improves overall utilization and support 10% more VoWLAN STAs than the traditional approach while the WLAN hotspot suffers from a heavy system loading such as 820 service requests.

4: CONCLUSIONS

In this study, a novel dynamic load balancing scheme that works together with a network-assist association policy was proposed for a VoWLAN system. The network-assist association policy tries to balance the loads between APs, and the dynamic load balancing scheme further optimizes the utilizations of APs in a WLAN hotspot in order to minimize the service rejection rate. Simulation results demonstrate that the proposed approach can improve the overall utilization of a WLAN hotspot by 6% to 10% while a WLAN hotspot system suffers from a heavy loading.

ACKNOWLEDGEMENT

The authors would like to thank the MediaTek-NCTU research center and ITRI-NCTU research center for financially supporting this research.

REFERENCES

[1]. D. Lenton, "Speaking of Wi-Fi," IEE Review, Vol. 49, Iss. 7, pp. 44 - 47, July 2003.

[2]. Moncef Elaoud el al., "Experimental VoIP Capacity Measurements for 802.11b WLANs," IEEE Consumer Communications and Networking Conference (CCNC), 2005.

[3]. Wei Wang, Soung Chang Liew, and Li, V.O.K. "Solutions to Performance Problems in VoIP over a 802.11 Wireless LAN," IEEE Transactions on Vehicular Technology, Vol. 54, Iss. 1, pp. 366-384, Jan. 2005.

[4]. Bo Li, Li Yin, K. Y. Michael Wong, and Si Wu, "An Efficient and Adaptive Bandwidth Allocation Scheme for Mobile Wireless Networks Using an On-Line Local Estimation Technique," Wireless Networks, Vol. 7, Iss. 2, March 2001.
[5]. Mark Davis, "Radio Resource Management: A Wireless Traffic Probe for Radio Resource Management and QoS Provisioning in IEEE 802.11 WLANs," 7th ACM International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems, October 2004.
[6]. Hakima Chaouchi and Anelise Munaretto "Adaptive QoS Management for IEEE 802.11 Future Wireless ISPs," Wireless Networks, Vol. 10, Iss. 4, July 2004.
[7]. IEEE 802.11 Standard, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification," 1999.
[8]. IEEE 802.11k Draft Standard, "Wireless Medium Access Control (MAC) and physical layer (PHY) specifications. Amendment 9: Radio Resource Measurement," 2005.
[9]. IEEE 802.11e Draft Standard, "Part 11: Wireless MAC and PHY Specifications: Amendment 7: MAC QoS," Draft 13, 2005.
[10]. Yigal Bejerano, Seung-Jae Han and Li (Erran) Li, "Fairness and Load Balancing in Wireless LANs Using Association Control," ACM MobiCom'04, 2004.