Stream-based Traffic Management for Mobile Devices in Wireless LAN

Zin Win Aye, Myat Thida Mon
University of Computer Studies, Yangon
zinucsl09@gmail.com, myattmon@gmail.com

Abstract

Nowadays, as a number of mobile devices are increasing rapidly, each user may use multiple traffics at the same time in wireless environment. The optimal traffic scheduling of mobile networks for multicasting and unicasting between various mixed mobile traffic is a challenging problem. Therefore, there are various researches concerning with traffic scheduling methods in order to have efficient transmission rate and to reduce traffic's bottleneck at Access Point (AP) for mixed traffic. In this paper, the various scheduling methods in wireless network are analyzed and the framework of a new scheduling method, a two-step stream-based traffic scheduling method is proposed to satisfy efficient some Quality of Services (QoS) parameters: throughput, fairness and delay. In the first step, stream based scheduling adaptively balances between unicast and multicast traffic and in the second step, compound scheduling combines Proportional Fair (PF) for non-real time traffic and Delay Related Scheduling for real time traffic.

1. Introduction

There has been a dramatic growth of modern mobile devices and the usage of multiple applications in the wireless networks such as Wi-Fi and cellular networks. Analysing mobile traffic is necessary to provide high quality mobile network services such as QoS management, traffic engineering, etc. In fact, several multi-flow applications consume a lot of mobile network resources. The transmission energy of multiples internet traffic such as multimedia traffic can be increased by filtering or grouping them. There are two important issues for internet traffic in wireless network to tackle: Quality of service and legal issues**.**

The wireless bandwidth, fairness and delay play an important role in the selection of the network parameters. For high throughput and low delay, traffic bottleneck at AP should be effectively reduced by better filtering and scheduling techniques in WLAN. Nowadays, users may access both real time and non-real time applications in wireless network. Moreover, this traffic may be multicast or unicast streams. An AP should effectively schedule for many jumble traffic from mobile users to solve QoS requirements.

QoS is considered as a main issue in delay sensitive and high throughput application as it requires packets to arrive at their destination on time with the least delay and maximum throughput. However, with the growing increases of internet users and their demands, today the network shares the bandwidth among web browsing, email, voice data, and video applications on the same network. Carrying different types of data which increases the load on the network may cause a bottleneck in the network. Due to the bottleneck and the competition among different kinds of traffic flow on the network at AP, packets of voice streaming may experience packet loss.

Therefore, many researches [3] focus on improving QoS with scheduling algorithms by addressing the resources sharing problem. Although the capacity of current technologies continues to grow, the requirements of mobile applications such as data storage, bandwidth, power allocation, fairness, and delay grow as well [1] [2] have been broadening for industries, Education, Entertainment, Insurance and Field Services. Some of the popular scheuling methods include Proportional Fair, Round Robin, Opportunistic and so on. Some emphasis on throughput, some are for fairness, some are for delay. Therefore, most scheduling and queuing algorithm will not solve QoS requirement for mixed traffic type of applications from various mobile users [3].

Employing each one of policies to handle only real time traffic or only non-real time traffic or employing policies in isolation scheduling to handle both multicast and unicast traffic or both real time and non-real time traffic has the following drawbacks:

- Under considering only one of policies in scheduling, while simply considering guaranteeing delay for RT traffic, the policy will weak the target of throughput maximization and while simply considering maximizing the system throughput in a fair way for NRT traffic, the policy will ignore delay constraint on RT packets. In case of high probability of QoS violation, enjoying multimedia services will not be possible.

- Under the isolated scheduling policies, although the system throughput and fairness seems to be maximized for a NRT application and packets delay seems to be minimized for a RT application, the system cannot solve the QoS violation and bottleneck of mixed traffic at AP to save the bandwidth by taking advantages of multicasting when the mobile users use mixed multicast and unicast application including real time and non-real time traffic in WLAN.

Therefore, in this paper we consider both real time and non real time traffic for multicast and unicast streams to get notonly maximum throughput and fairness butalso minimum delay, limiting the throughput of users close to the base station. It proposes two stage scheduling algorithms that can adapt to multicast and unicast streams in wireless networks using time division multiplexing for both real time and non real time traffic: stream based scheduling and compound scheduling. Moreover, the system efficiently classifies traffic mixed traffic not to delay in wireless environment.

2. Analysis of Scheduling Methods

Through the past decades, many schedulers were introduced to improve the performance of real-time applications or non-real time applications. These schedulers can be classified into groups; packet-based schedulers and frame-based schedulers, and so on. The new method should be efficient and fair, give a high throughput, be bandwidth-guaranteed and enhance the performance of the real time and non-real time applications over WLAN [7].

Therefore, we analyze many scheduling algorithms such as WFQ, CBWFQ, RR and SP to solve above requirements, each method has each function such as fairness, delay, and so on separately. This is a disadvantage of methods when the user use mixed traffic. There are some researches which consider only both real time and non real time traffic or some consider only between multicast and unicast applications. Therefore, the proposed algorithm based on stream for multicast and unicast applications solves the bottleneck issue in order to improve the QoS of mixture traffic and meet delay and fairness by considering RR with priority.

Firstly, Weighted Fair Queuing (WFQ) [12] classifies the traffic into different flows without requiring defined access lists to provide fair treatment for all types of traffic. This means that low-bandwidth traffic effectively has priority over high-bandwidth traffic because high-bandwidth traffic shares the transmission media in proportion to its assigned weight. However, WFQ has certain limitation that is not scalable if the flow amount increases considerably and not available on high-speed interfaces such as ATM interfaces.

Secondly, Class Based Weighted Fair Queuing (CBWFQ) [6] provides a solution to WFQ's limitations by providing user defined traffic classes instead of individual flows. The classes are defined by matching criteria on packet headers which include: Protocols (TCP, UDP, RTP, etc), IP addresses, TCP and UDP ports. Unlike standard WFQ, CBWFQ allows the system to define traffic classes and apply parameters, such as bandwidth and queue-limits, to these classes. CBWFQ queues are only held to their minimum bandwidth guarantee during periods of congestion, and can thus exceed this minimum when the bandwidth is available. Incoming packets are filtered and IP packets that meet the match criteria for a class are placed in their corresponding class queue. After defining a class according to its filter, the characteristics of the class such as bandwidth and maximum packet limit must be assigned. The CBWFQ uses the class weights to ensure that each class is serviced fairly. The key disadvantage with CBWFQ is that no mechanism exists to provide a strict-priority queue for real-time traffic, such as VoIP, Video conferencing and so on.

Next, Round Robin (RR) scheme is a choice to compensate the drawbacks of FCFS which also has low implementation complexity. Newly arrival packets queue up by flow such that each flow has its respective queue. The scheduler polls each flow queue in a cyclic order and serves a packet from any-empty buffer encountered; therefore, the RR scheme is also called flow-based RR scheme. RR scheduling is one of the oldest, simplest, fairest and most widely used scheduling algorithms, designed especially for time-sharing systems. They do offer greater fairness and better bandwidth utilization, and are of great interest. However, since to the lack of flexibility which is essential if certain flows are supported to be treated better than other ones [5] [8].

Lastly**,** Strict Priority (SP) [8] is another classical service discipline which assigns classes to each flow. Different classes may be associated to different QoS level and have different priority. The eligible packets associated to the flow with higher-priority classes are send ahead of the eligible packets associated to the flow with lower-priority classes. This is why it called "strict" since the eligible packets with lower-priority classes will never be sent before the eligible packets with higher-priority classes. Strict priority suffers from the same problem as that of FCFS, since a packet may also wait arbitrarily long time to be sent. Especially for the packets with lower-priority classes, they may be even starved by the packets with higher-priority classes.

In this paper, the first step scheduling for multicast and unicast queue considers not only traffic classification but also priority of real time traffic with RR to solve the drawback of CBWFQ and SP.

3. Proposed System

The system considers a network problem for multicast group and unicast group of mobile clients that operate in a heterogeneous wireless access network environment. It focuses on scheduling and queuing algorithm to implement a new filter for multicast and unicast traffic using specific criteria which is necessary for them. The approach targets the filtering of unicast and multicast traffic instead of real time and non-real time that occurs in some scheduling algorithms of wireless networks. Therefore, it presents a high level overview of three major parts: classification, the first step scheduling, and the second step scheduling for a solution to this problem with an optimal allocation of mobile users to multicast group and unicast group.

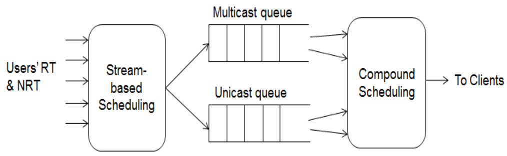

Figure 1.Overall system Architecture

Figure 1 is our proposed packet scheduling scheme operated in the AP. The classification part filters incoming traffic of mobile users into multicast and unicast traffic. Then, the stream based scheduling schedules the traffic of multicast queue and unicast queue fairly without much delay as a first step. Next, compound scheduling for mixed traffic of real time and non-real time is performed as a second step for some of QoS parameters: high throughput and good fairness and minimum delay.

3.1. Stream-based Scheduling for Multicast and Unicast Stream

As the number of mobile devices on the wireless networks has grown, not only the volume of unicast stream but also the volume of multicast stream from those devices has grown. Multicast is an efficient means of transmitting the same content to multiple receivers minimizing network resource usage being deployed over wireless environment. While multicast queue needs to send more packets than each node at a time, unicast queue send one packet each node at a time. In other words, the delay is unbalanced depending on the number of packets in multicast and unicast queue because resources are distributed unfairly. The unbalanced multicast and unicast queuing delay decrease the real time traffic capacity because delays need to meet the requirement for QoS of real time traffic.

Therefore, the system presents stream based scheduling to solve the problems as first step scheduling algorithm which is round robin scheduling with priority or round robin scheduling without priority for fairness of multicast and unicast queue without long delaying.

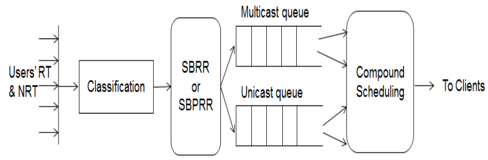

Figure 2.Detail system architecture

Figure 2 explains detail system architecture. The system classifies incoming mixed application according to their stream based different parameters in to two services group such as multicast or unicast queue. Both groups can have real time traffic and non-real time traffic due to mixed traffic of mobile users in WLAN simultaneously. For example, Video conferencing, Video streaming, Audio streaming and FTP and so on are belong to multicast group and web browsing, FTP, Video and BE are belong to unicast group. Then, it schedules each queue depending on priority scheduling methods such as SBRR or SBPRR. Next, compound scheduling calculates priority of real time and non-real time traffic in higher priority queue for QoS of both traffic types. There are assumptions for the system as follow:

Let assume that UB_r is bandwidth statistic of each user, AB is Available bandwidth of AP, MB_t is minimum bandwidth of each traffic, N be number of users in each queue, M be total number of users at AP, U be number of queues (same traffic), K is each user in the each queue.

$\sum_{i=1}^{U}\sum_{j=1}^{N} K \le M$ --------- eq (1)

$\sum_{r=1}^{M} UB_r \le AB$ ---------- eq (2)

$UB_r \ge MB_t$ --------- eq (3)

A. Classification

1: Available bandwidth at AP / Number of users= Average bandwidth of each user (ABW)

2: **If** ABW < minimum BW of each traffic
Reject incoming traffic
Else accept incoming traffic

3: **If** (UDP & same destination IP & port)
Send packets to multicast queue
Else send packets to unicast queue

There are several approaches to identify each packet's group. One suggestion is to classify each packet according to the service type of its protocol, port, source or destination addresses. Multicast works on UDP protocol and unicast is presented on TCP protocol. Likewise, real time traffic and non-real time traffic use UDP and TCP protocol respectively. The heterogeneous traffic from mobile users is filtered and put these into suitable queues according to their criteria. Each packet has observable features most of which are contained in the packet header.

Feature selection will affect the accuracy of traffic classification. It is sufficient to achieve high accuracy by using IP protocol, TCP/UDP ports, IP address as the features. The paper adopts these suggested features, since multicast traffic to be identified in our work are based on UDP protocol and IP address and port whereas unicast traffic is identified on TCP protocol [9].

B. First step scheduling

1: Count number of packets for both real time and non-real time traffic in each queue

$$Q_M = Q_{rt} + Q_{nrt}, \qquad Q_U = Q_{rt} + Q_{nrt}$$

2: Calculate the priority for each queue.

$$P = \frac{Q_M}{Q_U} \quad \text{if } Q_M \neq Q_U$$
$$P = 1 \quad \text{if } Q_M = Q_U$$

3: Schedule both multicast and unicast queue according to priority for in each round.

To achieve the fairness between the multicast and unicast traffic in an AP, when multicast and unicast queue have the same amount of traffic, the scheduling method needs to be able to schedule multicast packet and unicast packet sending all wireless clients using Stream based Round Robin (SBRR)within a given interval. Then, intuitively, the multicast queue needs to send N packets for many users in a given time while the unicast queue transmit one packet for each user in a given time interval. When multicast and unicast queue have not the same amount of traffic, the system use Stream based Priority Round Robin (SBPRR) method according to their number of packet in the each queue because it can adaptively change the number of the real or non-real time packets in each queue.

SBPRR would balance the multicast and unicast queuing delay in the case. In such a case, we need to consider the traffic volume of multicast and unicast in deciding the priority between the multicast and unicast queue. In order to consider such traffic volume changes, SBPRR uses the ratio of the number of packets in the queue (queue size) of the multicast and the number of packets in the queue of unicast as the priority of the multicast queue not to be queuing delay.

The dominant component of delay is the queuing delay considering that the transmission delay and the transmission overhead are very small. Furthermore, the transmission delay is the same in the queues assuming that they use the same transmission rate. Therefore, we need to balance the queuing delay of the multicast and unicast results in the balanced downlink delay. Thus, it is shown that SBPRR balances the queuing delay of the multicast and unicast traffic.

We can compute the queuing delay by multiplying the transmission time by the queue size according to little's law $D_{system} = Q_{system} / \mu_{system}$ [11]. Then, we can compute the queuing delay of the multicast D_M and the unicast D_U, multicast transmission rate μ_M, unicast transmission rate μ_U as follows:

$$D_M = Q_M \cdot \frac{1}{\mu_M} \qquad D_U = Q_U \cdot \frac{1}{\mu_U}$$

We consider the priority of the multicast P in two cases: When the queue size of multicast is greater than the queue size of unicast, $Q_M \neq Q_U$. Otherwise, when both queue size is equal, $Q_M = Q_U$. $\mu_M = P.\mu_U$ because the multicast transmits P packets when it gets a chance to transmit packets while unicast transmits only one packet. Thus, D_M can be rewritten as:

$$D_M = Q_M \cdot \frac{1}{P.\mu_U}$$

Then, we can get the optimal P value for balancing the delay of both queues as follows:

$$D_M = D_U$$

$$Q_M \cdot \frac{1}{P.\mu_U} = Q_U . \frac{1}{\mu_U}$$

Then,

$$P = \frac{Q_M}{Q_U} \quad \text{if } Q_M \neq Q_U$$
$$P = 1 \quad \text{if } Q_M = Q_U$$

Priority P is the ration of the queue size of multicast which is the number of multicast packets, and the queue size of unicast which is the number of unicast packets when both queue sizes is not equal.

For instance, if five wireless clients, from C1 to C5, each client sends two packets, and the multicast has six packets in its queue, then the queue size of unicast is four packets, and the priority of the queue P becomes (2 ~ 1.5 =6/4). Thus, in SBPRR, the multicast queue sends two packets in each time interval when the unicast queue acquires a chance to transmit a packet in each time interval. If we assume that each queue gets the same chance to transmit packet in each round, then the number of packets in the queue of the multicast and the number of packets in the queue of the unicast become two and one, respectively to send, and both of them become zero after the next transmission. Q_M / Q_U packets results one or less than one which means each queue gets the same chance to transmit packet in each round because both multicast and unicast have the same queuing delay.

Using this metric, the priority of the multicast changes adaptively when both queue's the traffic volume changes. When the amount of multicast traffic increases, the queue size of the multicast increases and the priority also increases to balance between queuing delay. When the queue size of unicast traffic increases, the priority of multicast traffic will decrease. Instead of using the number of packets, we could also use the packet size to compute the ratio between the multicast and unicast traffic volume. However, for our application, the overhead to transmit a voice packet is very large compared to the small voice data size. It was also confirmed that using the number of packets queued performs better than using the packet size. In order to balance between system delay and fair resource distribution the proposed scheduler utilizes the property of Round Robin (RR) with priority of the queue.

3.2. Second Step Scheduling for Real time and Non Real time traffic

Wireless systems tend to support heterogeneous traffic which includes real-time traffic (e.g., voice/video and gaming) and non- real time traffic (e.g., web browsing and file download). Different traffic classes have different QoS requirements, such as the delay bound of real-time traffic and the minimum average throughput requirement of non-real time traffic [10]. The users expect good service, but the wireless system wants to support more users and obtain a great system throughput.

So, how to design the scheduling algorithm to improve resource efficiency with all users' QoS ensured is not only an important problem, but also a complex problem. Packet scheduling can guarantee the different flows' QoS requirements by providing methods of resource allocation and multiplexing at the packet level. Analyzed paper proposed joint packet scheduling for real time and non-real time traffic by multiplying adjusted weight value for mixed traffic of mobile users.

The system proposes a scheduling algorithm to minimize user's average packet delay and to maximize the system throughput for multiuser cognitive networks, in which the user has different traffic models. The system schedules the multicast or unicast queue as first step scheduling in each round and then implements the compound scheduling method for mixed traffic of mobile users by multiplying the corresponded priority value with PF for non-real time traffic and DR for real time traffic. The detail procedure of the second step scheduling:

1: Calculate the priority for each traffic type.

$P_{rt} = \frac{Q_{rt}}{Q_M}$ for RT, $P_{nrt} = \frac{Q_{nrt}}{Q_M}$ for NRT

(OR)

$P_{rt} = \frac{Q_{rt}}{Q_U}$ for RT, $P_{nrt} = \frac{Q_{nrt}}{Q_U}$ for NRT

$P_{rt} + P_{nrt} = 1$

2: Schedule both real time and non-real time traffic according to priority.

For (schedule according to P in each round in the first step scheduling)

Schedule according to P_{rt} and P_{nrt} in each time interval

$$j = \max_{\substack{1 \le k \le K \\ 0 \le j \le 1}} \arg\{P_{nrt} * PF + P_{rt} * DR\}$$

The compound scheduling is determined to be summation of individual policies by multiplying each police with associated priority P_{rt} and priority P_{nrt} . $j = 0$ represents the RT stream and $j = 1$ represents the NRT stream. P_{rt} is the priority for first policy and P_{nrt} is the priority for the second policy [4].

3.2.1. Proportional Fair Scheduling

The well-known packet scheduling schemes for future wireless cellular networks are the maximum carrier-to-interference ratio (Max CIR), the proportional fair PF . For the efficient utilization of scarce radio resources under massive downlink traffic, throughput and fairness in wireless networks has been considered. PF tries to increase the degree of fairness among connections by selecting those with the largest relative channel quality where the relative channel quality is the ratio between the connection's current supportable data rate and its average throughput.

Therefore, the paper pays attention focusing on the PF scheme for non-real time traffic. It promises an attractive trade-off between the maximum average throughput and user fairness for the sake of completeness. Many of the algorithms rely on optimizing an objective function which ensures optimal performance [5].

PF favors users which has good channel conditions and high instantaneous data rate, in order to keep system throughput high. However, in the meantime it also considers user with bad channels since these user' low average rate $R(t)$ will increase their chance to be selected for next scheduling slot. This interesting tradeoff between fairness and performance results in wide acceptance and study of this scheme.

The PF scheduler a user with maximum priority by

$$PF = k^*(t) = \arg\max_{1 \le k \le K} \{D_{k,j}(t) / R_{k,j}(t)\}$$

$$R_{k,j}(t+1) = \begin{cases} (1-1/t_c)R_{k,j}(t) + 1/t_c D_{k,j}(t) & k = k^*(t) \\ (1-1/t_c)R_{k,j}(t) & k \neq k^*(t) \end{cases}$$

where $\arg\max$ denotes the argument of maximum, k is the index of user, PF is the selected user and K is the total number of user. $D_{k,j}(t)$ represents the instantaneous data rate that can be achieved by j stream of user k at time t and $R_{k,j}(t)$ represents the average data rate that j stream of user k perceived at time t. t_c is a time constant adjusted to maintain fairness over a pre-determined time horizon.

3.2.2. Delay related Scheduling

In our research, the popular delay-related QoS metric [4] will be employed. To define residual time metric, the ratio of a packet's time in the system to the delay threshold of the packet is given by

$$DR = d_{k,j} / d_{k,j}^{th}$$

where $d_{k,j}$ is the delay encountered by a packet at the head of the j^{th} stream of the k^{th} user. $d_{k,j}^{th}$ represents the delay threshold for packets at the j^{th} stream of the k^{th} user.

4. Conclusions

In this paper we proposed the framework of traffic management and scheduling to increase bandwidth usage and fairness without much delay for mobile user in wireless LAN to be able to solve bottleneck problem and scalability issues of traffic classification. The scheme has likely more performance in internet or data traffic with a high degree of compatibility with existing scheduling methods. The actual performance result of the system will be tested as an ongoing process. As a future work, we will investigate and implement the framework to be satisfied with QoS parameters with real wireless environment and consider congestion avoidance for mobile traffic.

References

[1] M. Alptekin Engin and Bulent Cavusoglu, IEEE communication letter, vol 15,no-11,November 2011,pg 1234-1237.

[2] S.S.Manvi and P. Venkataram, "Mobile agent based online bandwidth allocation scheme for multimedia communication", Electrical Communication Engineering Department, indian Institute of Science, Bangalore-560012 INDIA.

[3] K. Nisar, A. M. Said, H.Hasbullah, "A Voice Priority Queue (VPQ) Fair Scheduler for the VoIP over WLANs", Department of Computer & Information Sciences, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, 31750 Tronoh Perak, MALAYSIA, vol 3, no-2, Febuary 2011.

[4] H. Haci, H. Zhu and J. Wang, " Novel Scheduling for a Mixture of Real-time and Non-real-time Traffic", School of Engineering and Digital Arts, University of Kent, 2012.

[5] M. S. Mushtaq, Image, Signal and Intelligent Systems Laboratory-LISSI, Dept. of Networks and Telecoms, IUT C/V, University of Paris-Est Creteil (UPEC), France, A. Shahid and S. Fowler, Mobile Telecommunications, Dept. of Science and Tech, Link¨ oping University, Norrk¨ oping, Sweden, " QoS-Aware LTE Downlink Scheduler for VoIP with Power Saving".

[6] D. C. Vasiliadis, G. E. Rizos, C. Vassilakis, Department of Computer Science and Technology, University of Peloponnese, Tripolis, Greece, Technological Educational Institute of Epirus, Arta, Greece, " Class-Based Weighted Fair Queuing Scheduling on Quad-Priority Delta Networks".

[7] SUN Qiao-yun, LI Ze-min, CHAI Wen-yan, TIAN Hui, ZHANG Shu-guang, "A novel scheduling scheme for multiple traffic classes", The Journal of China Universities of Posts and Telecommunications, July 2010, 17(Suppl.): 91–94, www.sciencedirect.com/science/journal/10058885.

[8] T. Y. Tsai, Y. L.Chung and Z. Tsai, Institute for Information Industry, Graduate Institute of Communication Engineering, National Taiwan University, and Taipei, Taiwan, R.O.C, "Introduction to Packet Scheduling Algorithms for Communication Networks", www.intechopen.com.

[9] C. Faisstnauer, D. Schmalstieg, W. Purgathofer, Vienna University of Technology, "Priority Round-Robin Scheduling for Very Large Virtual Environments and Networked Games".

[10] C.Q. Dai, F. L. Liao, and Z. F. Zhang, Chongqing Key Lab of Mobile Communication Technology Chongqing University of Posts and Telecommunications Chongqing China, "Packet Scheduling for Downlink OFDMA Wireless Systems with Heterogeneous Traffic", Journal of Network, vol-7, no-6, June 2012.

[11] S. S. H. Schulzrinne, Department of Computer Science Columbia University, "Towards the Quality of service for VoIP traffic in IEEE 802.11 Wireless Networks", 2008.

[12] H. J. Haas, "Queuing Methods From FIFO to CB-WFQ/PQ", May 2005, www.perlhel.at.