Bandwidth Management for Mixed Unicast and Multicast Multimedia Flows with Perception Based QoS Differentiation

Guowang Miao, Zhisheng Niu
State Key Lab on Microwave and Digital Communications
Department of Electronic Engineering, Tsinghua University, Beijing, 100084, P.R. China,
Telephone: +86-10-62794328 Fax: +86-10-62773634

***Abstract*— Efficient bandwidth allocation is among the most important open issues in network traffic control aiming to offer a QoS assurance, especially when unicast and multicast multimedia flows coexist. With new QoS concepts: user/flow satisfaction, this paper presents a user perception oriented QoS model, and a generalized bandwidth allocation method which can actively adapt to the dynamics of the network and has global convergence capability for an optimal solution of bandwidth allocation. Our method effectively exploits all available bandwidth resources, and can provide effective QoS differentiated services when different types of unicast and multicast multimedia streams coexist. The proposed method has great flexibility that QoS differentiation can be guaranteed in both heavy and light load conditions. For heavy load cases, the allocation results will satisfy all users as much as possible. And for light load cases, all redundant bandwidth will be fully exploited while distinct QoS differentiation is kept. Based on our approach, a soft system capacity similar to the one in CDMA networks will be provided.**

I. Introduction

While there are other criteria to measure QoS like delay or jitter, bandwidth is of interest to the largest number of applications. Bandwidth allocation becomes more difficult, especially for QoS assurance, in future multi-service networks which consist of different kinds of unicast and multicast multimedia flows. There are two prevailing classes of bandwidth allocation: static allocation and dynamic allocation. Since static bandwidth allocation policies lack adaptive mechanism to combat the dynamics in network, many studies have already focused on dynamic bandwidth allocation schemes. Lesiak et al. [2] given a description of the threshold-based algorithm, where a bandwidth modification request is requested wherever a threshold is crossed. This is a most basic dynamic allocation method, which does not consider either the flow characteristics or QoS assurance. [3] provides a harmonic proportional allocation scheme to guarantee proportional sharing between classes with different priorities. [1] considers that applications can tolerate a certain QoS degradation and exploits an allocation with multilevel degradable quality. Based on differential pricing, [4] analyzes the fairness and truthfulness properties from a game-theoretic perspective and gets the allocation result from Nash equilibrium to satisfy the max-min fairness. However, these papers([3], [1], [4]) focus too much on QoS assurance, and the traffic flows competing for bandwidth allocation are modelled too simple and contain one type of applications only, e.g. elastic applications only in [4]. Other researches based on utility concept like [5], [6] give good modelling of multimedia traffics, and give bandwidth allocation when different kinds of multimedia traffics coexist. However, what they are most interested in is fairness between different flows, which means they are lack of effective QoS differentiation ability. Besides, they do not consider the coexistence of unicast and multicast multimedia flows, each of which may contain different types of applications.

However, the coexistence of unicast and multicast multimedia flows will occupy more and more part of future communications. This motivate us to develop the bandwidth allocation in this paper. Based on new concepts: user and flow satisfaction factor, a user perception oriented QoS model is introduced for unicast and multicast multimedia traffics and a generalized bandwidth allocation scheme which actively adapts to the dynamics of network is developed. Our scheme also considers the characteristics of incoming stream, which might greatly injure the exploitation of available bandwidth and have rarely been considered in existing researches. The QoS policy has great flexibility that effective QoS differentiation can be guaranteed in both heavy and light load conditions for both unicast and multicast multimedia flows. For heavy load cases, the allocation results will satisfy all users as much as possible. for light load cases, all redundant bandwidth will be exploited while distinct QoS differentiation is kept.

II. Policy Modelling and Dynamic Bandwidth Allocation Optimization

The bandwidth allocation policies will be executed at each single link, which contains unicast and/or multicast traffic flows received by a user or a group of users. The reallocation will be executed when some existing flow has finished its transmission or a new connection needs to be established.

A. Network Dynamic Characteristics

Owing to the varying load of both network links and nodes, the end-to-end bandwidth between the ISP and current routing node varies between different connections or even changes rapidly during one connection time. Hence the varying data rate of incoming stream may be lower than the data rate request the flow carries and lower than the bandwidth allocated by the current node, which causes the inefficient utilization. Measuring end-to-end bandwidth is an interesting and tough problem frequently discussed in recent years, and many policies and techniques have been proposed ([7], [8], etc.). The bandwidth allocated to each flow should be no more than the available bandwidth from ISP to the current node: $R_i \leq B_i$, where R_i is the bandwidth assigned to flow i, and B_i is the available end to end bandwidth from the ISP to the current node for flow i.

B. User Satisfaction factor

Users can not always be satisfied. The ***user satisfaction factor*** is defined to give some yards to measure the services every customer acquires:

$$\gamma_i = f_i(S_i), \tag{1}$$

where S_i is the service user i receives. $f_i(S_i)$ is the satisfying function for calculating contentment of user i receiving service S_i, and γ_i is the user satisfaction factor representing how much user i is satisfied with the obtained service. γ_i is scaled from 0 to 1, and the bigger γ_i is, the more satisfied user i will be.

Different satisfying functions should be designed for different multimedia applications according to their characteristics. Real-time applications like video conference, voice over IP, etc, can be tolerant of a certain packet drops, but the perception of the application quality will degrade sharply once the allocated bandwidth is smaller than the minimum encoding rate. However, the over provision of bandwidth has only slight improvement of the user's perception of application quality. Meanwhile, rate-adaptive applications like file transfer, email, etc, show much better adaptive capability. Denote R for the allocated bandwidth, γ_0 for the basic satisfaction factor when the user is assigned with the requested bandwidth C. We design satisfying functions for real-time and rate-adaptive applications in eq.(2) and eq.(3) respectively.

$$\gamma = f(R) = \begin{cases} 0 & 0<R\leq\gamma_0 C \\ \gamma_0 - \gamma_0\sqrt{1-\frac{(R-\gamma_0 C)^2}{(1-\gamma_0)^2C^2}} & \gamma_0 C<R\leq C \\ \gamma_0 + (1-\gamma_0)\sqrt{1-\frac{(R-(2-\gamma_0)C)^2}{(1-\gamma_0)^2C^2}} & C<R\leq(2-\gamma_0)C \\ 1 & R>(2-\gamma_0)C \end{cases}, \tag{2}$$

$$\gamma = f(R) = \begin{cases} \gamma_0\frac{(R)^2}{(C)^2} & 0 < R \leq C \\ 1-\frac{(1-\gamma_0)C}{R} & R > C \end{cases} \tag{3}$$

Fig.(1) shows a critical real-time application satisfying function, in which $C = 1.024Mbps$ and $\gamma_0 = 0.9$. Fig.(2) shows a rate-adaptive application where $C = 1.024Mbps$ and $\gamma_0 = 0.5$. As shown in Fig.(1) and Fig.(2), the designed satisfying functions give proper descriptions of the relationship between the user perceived quality and allocated bandwidth.

C. Satisfaction Oriented Allocation Method (SOAM)

To satisfy all users' requirement as good as possible is our purpose. Flow i is marked with priority P_i, and $1 \leq P_i \leq K, K > 1$. According to the flow priorities:

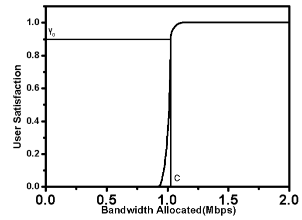

Fig. 1. Real-time Applications

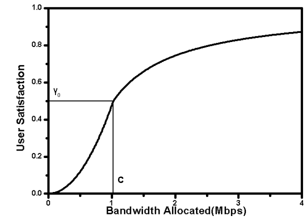

Fig. 2. Rate-adaptive Applications

$$E(\gamma_i) = g(P_i), \tag{4}$$

where $E(\gamma_i)$ is the expectation of the satisfaction factor for user i, and $g(\bullet)$ is the priority-satisfaction function which reveals the relationship between priority and satisfaction factor. $g(\bullet)$ is assumed in our policy as:

$$E(\gamma_i) = g(P_i) = \gamma_0 + (P_i - 1)\frac{\gamma_m - \gamma_0}{K-1} \tag{5}$$

where $\gamma_m = max\{f_i(S), i = 1, \ldots, N\}$, in which S is the maximum capability system can supply, and N is the number of active flows. According to the available resources, more or less services can be provided to fully exploit the system resources. The user satisfaction factor can be calculated based on the allocated services, and according to priority-satisfaction relationship, the actual priority of service user i receives is:

$$P_i^{'} = g^{-1}(\gamma_i) = (K-1)\frac{\gamma_i - \gamma_0}{\gamma_m - \gamma_0} + 1. \tag{6}$$

The objective is twofold: first, we want to increase satisfaction of all users; second, we want to assure fairness and avoid that users with higher priority are too greedy. Besides, when unicast and multicast flows coexist, they should be treated differently. As shown in Fig.3, an unicast flow received by R_{11} and a multicast flow received by $R_{2i}(i = 1, 2, 3)$ coexist. The service provided for flow 2 will be perceived by many users while flow 1 by 1 user only. Hence we should consider more for flow 2 even these two flows have the same priority and carry the same kind of applications. However, we could not simply consider the sum of all receivers' satisfaction as the metric of flow bandwidth allocation. Because a multicast flow with lots of receivers will probably occupy all the bandwidth to achieve a large total satisfaction, which does not make too much sense and would be unfair to other flows.

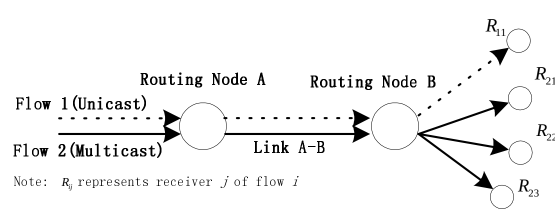

Fig. 3. Coexistence of Unicast and Multicast Flows

Hence we need to establish a flow QoS metric, which incorporates the characteristics of both unicast and multicast.

The ***flow satisfaction factor*** for each flow is defined for this purpose:

$$\alpha_i = W(U_i)P_iP_i', \tag{7}$$

where α_i is the flow satisfaction factor, U_i is the number of users that will receive this flow and $U_i = 1$ for unicast flows, $U_i > 1$ for multicast flows. $W(U_i)$ is the multicast weight factor. $W(U_i)$ is an increasing function, and should converge to a constant when $U_i \to \infty$. A typical $W(U_i)$ is designed as

$$W(U_i) = K - \frac{K-1}{(U_i)^{log_2 \frac{K-1}{K-1.5}}}, \tag{8}$$

which is K when U_i converges to infinite.

Our purpose is to maximize the satisfaction factor of all flows, defined as:

$$\mathbb{TS} = \sum_{i=1}^{N} \alpha_i, \tag{9}$$

where N is the number of active flows.

Bandwidth allocated to critical real-time applications should not be reallocated. Since users receiving these applications would be even unhappier if their applications are suddenly interrupted for the reallocation. Assume that the total bandwidth of the current node is R, the bandwidth that has been occupied by critical real-time applications is R_c. The remaining bandwidth will be reallocated by the other flows or new connections. Assume that N flows either unicast or multicast will share the available bandwidth. Considering the varying data rate of incoming streams, the objective is:

$$\begin{aligned} &\mathbb{TS}_{(1)} = \max\left\{ \sum_{i=1}^{N} \left[W(U_i)P_i g^{-1}(f_i(R_i)) \right] \right\}, \\ &s.t. \begin{cases} R_i \le B_i & i = 1\ldots,N \\ \sum_{i=1}^{N} R_i + R_c \le R & \\ R_i \ge 0 & i = 1\ldots,N \end{cases}, \end{aligned} \tag{10}$$

where B_i is the data rate of incoming stream of flow i.

Without considering the varying data rate of incoming streams, the objective will be:

$$\begin{aligned} &\mathbb{TS}_{(2)} = \max\left\{ \sum_{i=1}^{N} \left[W(U_i)P_i g^{-1}(f_i(R_i)) \right] \right\}, \\ &s.t. \begin{cases} \sum_{i=1}^{N} R_i + R_c \le R & \\ R_i \ge 0 & i = 1\ldots,N \end{cases}, \end{aligned} \tag{11}$$

The bandwidth allocation is nonlinear programming which contains nonlinear objective function and simple linear constraints. Many techniques have been proposed for solving constrained optimization problems([9], [10]). Considering the specific form of our optimization objective and its constraints, the proposed allocation algorithms will be based on Rosen's gradient projection method([11]), which is based on successive projections of the gradients on the subspace tangent to the active constraints, and has inexpensive operations when the constraints are simple. The Kuhn-Tucker conditions will be used as a termination criteria. The complete satisfaction oriented allocation method(SOAM) is shown in TABLE I(for detailed theoretical analysis, please refer to [11]).

TABLE I

SATISFACTION ORIENTED ALLOCATION METHOD (SOAM)

1.***Initialization:***	Set the initial feasible point (allocation) $\mathcal{R}^{(0)}$, k=0
2.***Projection:***	Terminate if $\mathcal{R}^{(k)}$ is K-T point or iterations have exceeded the limit; otherwise,compute the feasible direction $d^{(k)}$
3.***line search:***	Search the optimal step size by line search and get $\mathcal{R}^{(k+1)}$
4.***update:***	Set $k \leftarrow k+1$,and go to 2.

III. PERFORMANCE ANALYSIS OF SOAM

A. The Global Convergence

Theorem 1: If all satisfying functions of each user are designed to be: $f(S) \in C^1$, i.e., $f : \Omega \to R^1$, in which Ω is the feasible domain, is continuously differentiable, the Satisfaction Oriented Bandwidth Allocation Method without iteration limit will either stop at a Kuhn-Tucker point or generate an infinite sequence whose cluster points are all Kuhn-Tucker points.

Proof: Consider the problem $minf(x) : A^Tx \ge b, x \in R^m$. Based on the following two assumptions:

Assumption 1:$f(x) \in C^1$, i.e., $f : \Omega \to R^1$, is continuously differentiable;

Assumption 2:For any feasible point x, $\{a_i \mid a_j^T x = b_j\}$, whose elements are column vectors of a submatrix of A, is a set of independent vectors,

[13] has proved that under the different disposal of new feasible direction, the Rosen's projection algorithm either stops at a Kuhn-Tucker point or generates an infinite sequence x^k in which every cluster point is a Kuhn-Tucker point. In SOAM, with the continuously differentiable design of the satisfying functions, the target function is also continuously differentiable. Besides, the coefficient matrix of the constraints contains a set of independent vectors. Hence, we have the conclusion that without iteration limit, SOAM either stops at a Kuhn-Tucker point or generate an infinite sequence whose cluster points are all Kuhn-Tucker points.

The two satisfying functions designed for critical real-time applications and rate-adaptive applications are continuously differentiable. Hence according to theorem 1, the bandwidth allocation method has global convergence capability.

B. QoS Differentiation Capability and Design of Satisfying Function

Definition: Assume the bandwidth allocation is $\mathcal{R} = [R_1, \cdots, R_N]^T$, total bandwidth $R = \sum_{i=1}^{N} R_i$. With an infinite small increase of total bandwidth: $R \to R+\Delta R$, define the bandwidth competition capability of flow i at allocation point $\mathcal{R}$ as:

$$C_i = \lim_{\Delta R \to 0} \frac{\Delta \mathbb{TS}_i}{\Delta R}, \tag{12}$$

where $\Delta\mathbb{TS}_i = \mathbb{TS}' - \mathbb{TS} = G(\mathcal{R}) - G(\mathcal{R}_i')$, in which $\mathcal{R}_i' = [R_1, \cdots, R_i + \Delta R, \cdots, R_N]^T$.

This definition is reasonable, since the bigger C_i is, the more total satisfaction will be achieved if the increased bandwidth ΔR is allocated to flow i. Our allocation is total satisfaction maximization oriented, and the increased bandwidth will be allocated first to flows with larger C_i to get larger total satisfaction. Hence C_i directly reflects the bandwidth competition capability of flow i at allocation point $\mathcal{R}$.

Theorem 2: At $\mathcal{R} = [R_1, \cdots, R_N]^T$, the bandwidth competition capability of flow i is proportional to the slope of satisfying function at point R_i of flow i in SOAM.

Proof: slope of satisfying function at R_i is: $k_i = \frac{\partial f_i(\mathcal{R}_i)}{\partial R_i}$. The bandwidth competition capability of flow i at $\mathcal{R}$ is:

$$\begin{aligned} C_i &= \lim_{\Delta R \to 0} \frac{\Delta \mathbb{TS}_i}{\Delta R} = \frac{\partial(-G(\mathcal{R}_i))}{\partial R_i} \\ &= W(U_i) P_i \frac{\partial(f_i(R_i))}{\partial R_i} / \frac{\partial(g(P_i^{'}))}{\partial P_i^{'}} \cdot \frac{C_i}{k_i} \\ &= W(U_i) P_i / \frac{\partial(g(P_i^{'}))}{\partial P_i^{'}}. \end{aligned} \tag{13}$$

Hence C_i is proportional to the slope of satisfying function at point R_i of flow i.

Based on Theorem 2, the satisfying functions for different kinds of applications can be designed. In order to protect the point that is sensitive to bandwidth requirement, the satisfaction function should have large slope at the corresponding bandwidth point, and the more sensitive, the larger the slope should be. For example, a video with adaptive codec and modulation may generate a data stream with different possible data rates, hence this application have several sensitive bandwidth points. The satisfying function should be designed to slant deep at these bandwidth points.

Theorem 3: At $\mathcal{R} = [R_1, \cdots, R_N]^T$, the bandwidth competition capability of flow i is proportional to the flow priority P_i in SOAM.

Theorem 4: At allocation point $\mathcal{R} = [R_1, \cdots, R_N]^T$, the bandwidth competition capability of flow i is proportional to the multicast weight factor when unicast and multicast flows coexist in SOAM.

Since $C_i = \lim_{\Delta R \to 0} \frac{\Delta \mathbb{TS}_i}{\Delta R} = W(U_i) P_i \frac{\partial(f_i(R_i))}{\partial R_i} / \frac{\partial(g(P_i^{'}))}{\partial P_i^{'}}$, it is easy to see that both Theorem 3 and 4 are correct.

From *Theorem* 3 and 4, we know that SOAM provides QoS differentiation with different bandwidth competition capabilities when unicast and multicast multimedia flows with different priorities coexist.

According to the characteristics of bandwidth requirement of different kinds of applications, the corresponding satisfying functions can be designed based on *Theorem* 2, 3 and 4.

C. System Capacity Analysis of SOAM and Call Admission Control for Basic QoS Assurance

Based on SOBAM, there's no upper bound of the number of flows that can be connected simultaneously. This happens especially when most connected flows carry rate-adaptive applications, which will get a portion of the reallocated bandwidth when a new connection request arrives. Hence for rate-adaptive applications, SOBAM has soft capacity similar to the one in CDMA systems. Any new flow can compete with the other existing flows for connection any time, which will cause the bandwidth reallocation and make users receiving other flows less satisfied.

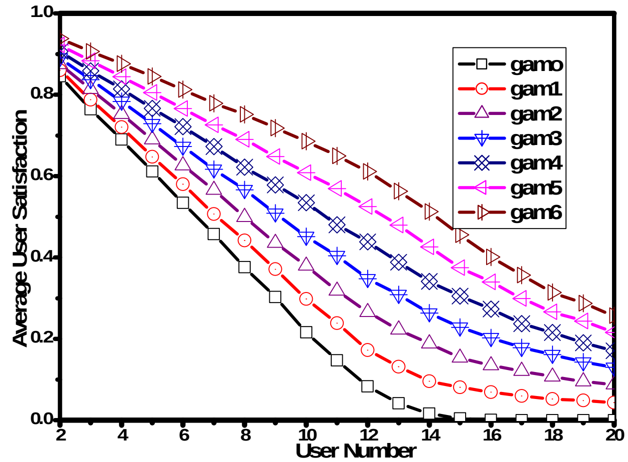

Fig. 4. Relationship between Average User Satisfaction and System Capacity

Fig4 gives the simulation result, and shows the relationship between average user satisfaction and system capacity(measured by online user number) under different settings(basic satisfaction factor γ_0). We assume that the total available router bandwidth is 4096kbps, and all connections are unicast flows with the same priority, and are requesting ftp services with basic data rate 384kbps. Seven curves are plotted to illustrate the simulation results, and are marked by 'gam$i$',$i = 0, \ldots, 6$ to show the capacity-satisfaction relationship when γ_0 is configured to be $0, 0.1, \ldots, 0.6$ respectively. As shown in Fig4, the average satisfaction decreases in a quasi linear way with the increase of the flows served.

However, users should not be too disappointed when they are receiving a certain services. The router may choose to reject new connections rather than provide worse services to all users when too many flows already exist. And thus a basic service level can be assured. Hence, some call admission control(CAC) mechanisms are needed before new connections are accepted. The following steps are designed to make an admission decision and the bandwidth allocation after the decision:

- Step1 Update the router condition, and get R,R_c. Get the parameters of the new arriving flow and the existing flows that will take part in the bandwidth reallocation.
- Step2 Reallocate the bandwidth beforehand according to SOBAM, and get the allocation result: $R_1, R_2, \cdots, R_s, R_n$, where the first s allocations are for the existing flows joining in the bandwidth reallocation and R_n is for the new arriving flow.
- Step3 Compute the user satisfaction factor according to (1), and then get the actual priority each user will receive during the reallocation according to (6):$P_1^{'}, \cdots, P_s^{'}, P_n^{'}$.
- Step4 If $\min_{i \in \{1,2,\cdots,s,n\}} \{\frac{P_i^{'}}{P_i}\} \geq \alpha_0$, in which α_0 is the basic flow satisfaction factor configured by the

router, then accept the new connection and execute the bandwidths reallocation. Otherwise, reject the new connection and keep the original router state.

It can be seen from the steps above that the CAC policy designed is to guarantee a basic satisfaction level of each flow corresponding to the flow priority.

IV. SIMULATION PERFORMANCE

Apply our algorithm to a router for CAC and bandwidth allocation. Assume that the arrivals of new connections are Poisson processes with gaussian distributed data rate requirement. Even though recent studies found that the packet arrival process in the Internet switches/routers is not Poisson, the network traces also show that the user-generated connection requests, such as Telnet or FTP connection requests, can still be modelled as a Poisson process. Since we mainly focus on the admission control and bandwidth allocation, which are the connection-level resource management, the Poisson process is a good approximation for our purpose. For mathematical tractability, we assume the lifetime of each connection to be exponentially distributed. Data flows are classified into 5 priorities. The basic satisfaction factor is 0.2 for unicast data flows, and 0.9 for multicast data flows.

First, consider the coexistence of unicast and multicast flows. A multicast flow containing DTV program and three unicast flows generated from FTP content downloading will compete for the available 10Mbps router bandwidth. The multicast flow has priority 2, and has 20 receivers. Thanks to the compression of original videos and the traffic smoothing techniques used in the MPEG-2 encoder([12]), the bandwidth required for transmitting DTV programs varies slightly from time to time. Hence, the multicast flow is assumed to be near CBR stream having Gaussian distributed bandwidth requests with mean 6Mbps and variance 100Kbps. The three unicast flows, marked with priorities 4,3 and 1, have basic data rates requests 256Kbps, 384Kbps and 1.024Mbps respectively. The Internet link bandwidth between the router and the ISP servers for the three unicast flows are assumed to be gaussian distributed with mean 2.3Mbps, 1.3Mbps, 1.9Mbps and variance 0.3Mbps, 0.2Mbps and 0.5Mbps respectively. We collect the simulation results for a random 100 reallocation times when these four flows coexists. Fig.5 shows how these flows are satisfied during each bandwidth reallocation for illustration, in which the four curves are marked by "M_P2","U_P4","U_P3","U_P1", corresponding to the flows above. The simulation results prove that the SOBAM algorithm can effectively supply services to users with different satisfaction according to their QoS level. The mean satisfaction of users receiving the four flows are 0.94,0.87,0.75 and 0.54 respectively measured by satisfaction factor.

Fig.6 presents the simulation results examining the influences of incoming stream characteristics to the utilization efficiency of current router bandwidth. Two unicast streams A and B with priorities 1 and 3 are assumed to compete for the remaining $4Mbps$ router bandwidth with data rate requirement 1.536Mbps and 2.048Mbps respectively. Meanwhile The link between the router and remote ISP for stream B is supposed to be very well, and is gaussian distributed with mean 3.072Mbps and variance 500Kbps. The bandwidth between the router and ISP for stream A is set to be gaussian distributed with variance 200kbps, and the mean value is set to increase from 0.2Mbps to 6Mbps by step 200Kbps(x coordination in Fig.6). Both the allocation algorithms with consideration of the Internet bandwidth(optimization(10)) and without(optimization(11)) are simulated independently, and the results are illustrated by curve A and B respectively to show the corresponding router bandwidth utilization efficiency. For each data point in Fig.6, we get statistics over 5000 reallocating times. As shown in Fig.6, the consideration of the varying Internet characteristics would greatly improve the router resource utilization efficiency, especially for bad Internet link conditions. The convergence of curve A and B shows that for good Internet conditions between the ISP and current node, allocations both with and without considering the incoming stream characteristics have nearly the same bandwidth utilization efficiency at current node.

Thanks to the design of satisfying functions, SOBAM has great flexibility that it can adapt the allocation results to the variation of each contained flow and the available bandwidth conditions of the router. In other words, SOBAM can give good allocation results with different router load conditions. Define the router load as: $\rho = \frac{\sum C_i}{R}$, in which C_i is the data rate request of user i, and R is the total available reallocation bandwidth. We give simulations over different router load situations. A multicast flow containing video conference and four unicast flows generated from FTP content downloading will compete for the bandwidth. The multicast flow, with 15 receivers, has priority 2, and has Gaussian distributed bandwidth requests with mean value 512kbps and variance 50Kbps. The four unicast flows, marked with priorities 1,2,4 and 5, have the same basic data rates requests 1024kbps. The Internet link bandwidth between the router and the ISP servers for each flow are assumed to be good enough. The available bandwidth for reallocation in the router is set to increase from 2Mbps to 30Mbps by step 1Mbps, which means a decreasing router load: $\rho_1 = \frac{0.512+1.024*4}{2} = \frac{4.608}{2}$ $\rho_2 = \frac{0.512+1.024*4}{3} = \frac{4.608}{3}, \ldots, \rho_1 = \frac{0.512+1.024*4}{30} = \frac{4.608}{30}$. Fig.7 presents the simulation result, in which we get statistics over 1000 reallocating times for each data point. x coordination is marked with the inverse of router load:$\frac{1}{\rho}$. And the curves plotted, legended by "M_P2" for the multicast flow and "U_Pi" for the unicast flow with priority i, show the average user satisfaction for each flow when the router has different load(from heavy to light). As shown in Fig.7, with the increase of router available bandwidth, all users will get more and more satisfied, and the allocation will make full use of all available bandwidth. Besides, it can be seen that for different traffic load cases, the satisfaction(QoS metric) supplied to different flows are effectively differentiated. The router utilization ratio has always been 100% during all the simulations, which shows that the available router resources have been plentifully exploited.

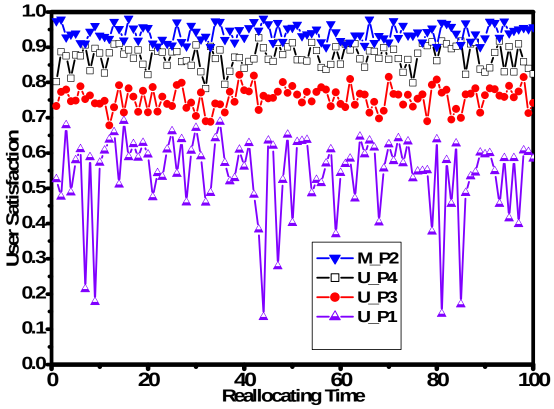

Fig. 5. User Satisfaction Trace

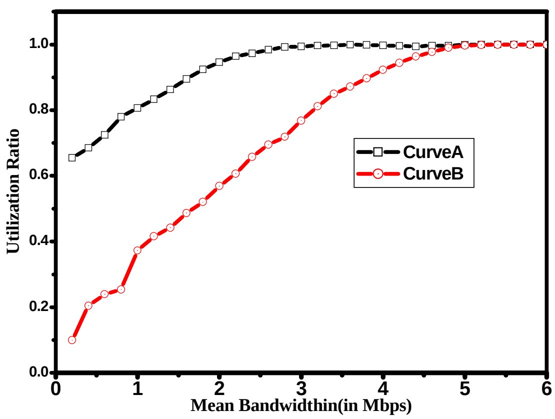

Fig. 6. Router Utilization Efficiency Under Different Network Conditions

V. Conclusions

With new QoS concepts: user/flow satisfaction, this paper has introduced a user oriented QoS model and a generalized bandwidth allocation scheme which actively adapts to the dynamics of network. The kind of mechanisms developed is a generalized scheme suitable for deployment at various routing points in the communication traffic which may consist of both unicast and multicast traffics. Our approach allows a service provider to differentiate the services between different types of customers based on their priority and service type. The QoS policy for multimedia traffic has great flexibility that effective QoS differentiation can be guaranteed in both heavy and light load conditions. For heavy load cases, the valuable bandwidth will be allocated to satisfy the users as much as possible. And for light load cases, all the redundant bandwidth will be fully exploited while distinct QoS differentiation is kept. The allocation scheme SOAM contains two allocation methods: the first one considers the incoming stream characteristics and can be used for bad Internet conditions to make efficient use of the current node bandwidth, while the second one doesn't consider the incoming stream characteristics and can be used for good Internet conditions for execution simplicity. The simulation results powerfully show that SOAM give good QoS differentiation to different users and make plentiful utilization

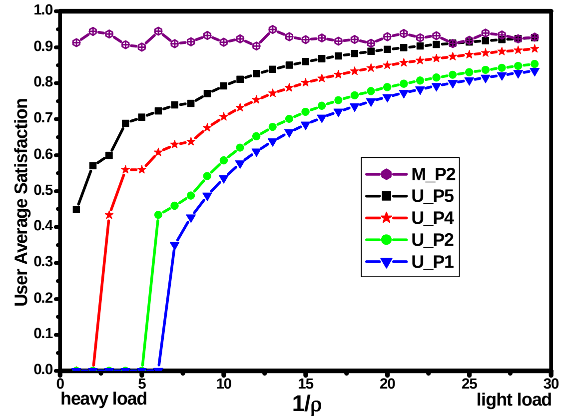

Fig. 7. QoS Ability of SOAM for Different Load Situations

of all available bandwidth. With SOAM, the system can provide a soft system capacity which is similar to the one in CDMA networks.

References

[1] Chun-Ting Chou; Shin, K.G. *Analysis of adaptive bandwidth allocation in wireless networks with multilevel degradable quality of service*. Mobile Computing, IEEE Transactions on Volume 3, Issue 1, Jan-Feb 2004 Page(s):5 - 17

[2] LESIAK, J., ROBLES-ROJI, F., and TOHME, S. *Bandwidth inaiiagemenl for LAN interconnection: IP over ATM*. IFIP/ICCC conference on Information network and data communication, 1996

[3] Zhou, X.; Xu, C.-Z. *Harmonic proportional bandwidth allocation and scheduling for service differentiation on streaming servers* . Parallel and Distributed Systems, IEEE Transactions on Volume 15, Issue 9, Sept. 2004 Page(s):835 - 848

[4] Haikel Y, Ravi R. M. and Catherine R. *A Game Theoretic Framework for Bandwidth Allocation and Pricing in Broadband Networks* . IEEE/ACM TRANSACTIONS ON NETWORKING, VOL. 8, NO. 5, OCTOBER 2000

[5] Zhiruo Cao; Zegura, E.W., *Utility max-min: an application-oriented bandwidth allocation scheme* . INFOCOM '99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE Volume 2, 21-25 March 1999 Page(s):793 - 801 vol.2

[6] Tobias Harks, *Utility Proportional Fair Bandwidth Allocation: An Optimization Oriented Approach* . QoS-IP 2005, LNCS 3375, pp. 61C74, 2005. Springer-Verlag Berlin Heidelberg 2005

[7] M.Jain, C. Dovorlis, *End-to-end Available Bandwidth:Measurement Methodology, Dynamics and Relation with TCP Throughput*. ACM SIGCOMM'02, Pittsburgh, Pennsylvania, USA, Aug.2002.

[8] B.Melander, M. Björkman, P. Gunningberg *A New End-to-End Probing and Analysis Method for Estimating Bandwidth Bottlenecks*. Proc. IEEE GLOBECOM, San Francisco, Nov.2000.

[9] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, New York, 1982

[10] R. Fletcher, *Practical Methods of Optimization*. 2nd ed., John Wiley &Sons, Inc., New York, 1987

[11] J.B.Rosen, *The gradient projection method for nonlinear programming,part I:linear constraints*. Journal of the Society for Industrial and Applied Mathematics 8, 181-217, 1960

[12] Conti, M.; Gregori, E., *Global Telecommunications Conference, 1997*. GLOBECOM '97., IEEE , Volume: 1 , 3-8 Nov. 1997 Pages:443 - 447 vol.1 .

[13] Ding-Zhu Du, Xiang-Sun Zhang, *A Convergence Theorem of Rosen's Gradient Projection Method*. Mathetical Programming, vol 36(1986),135-144, Amsterdam, North-Holland.