

Dynamic Traffic Management Focusing on User's Satisfaction in All-IP Mobile Environment

Nobutoshi Abe¹, Norihiko Shinomiya²,
and Yoshimi Teshigawara¹

¹ Graduate School of Engineering

² Department of Engineering

Soka University

1-236 Tangi-cho, Hachioji-shi, Tokyo JAPAN

{nabe,shinomi,teshiga}@soka.ac.jp

Abstract

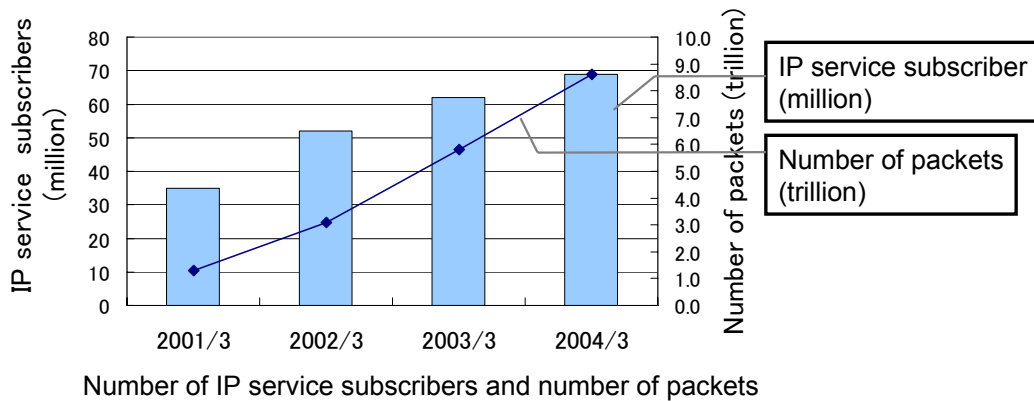
The next-generation All-IP mobile communication environment requires an efficient traffic management scheme to achieve both establishing simultaneous connectivity for many users and guaranteeing QoS requirements of various application services. Connectivity and QoS guarantee such as bandwidth have a trade-off relationship. There have already been some traffic management schemes which have the goal to maximize the throughput in a base station or control the traffic by considering either one side of the trade-off or the other. Users, however, are not necessarily satisfied with a provided service, even if employing these schemes. It is essential to balance both sides of the trade-off, because of a crucial issue to provide services that satisfy users in a competitive wireless telecommunication market. This paper proposes a flexible traffic management scheme considering the user's satisfaction rate (USR) derived from a session request blocking rate for the simultaneous connectivity and an acquired bandwidth specific to a streaming service for the QoS guarantee. Our simulation analysis demonstrates that the proposed scheme improves the minimum value of the USR compared with other conventional systems. In addition, it indicates the achievement of balancing both sides of the trade-off.

Keywords: All-IP mobile network, User Satisfaction Rate, QoS,
Dynamic Traffic Management, AHP

Contact Person: Nobutoshi Abe (nabe@soka.ac.jp)

Introduction

- Requirements for All-IP mobile network
 - Large capacity system for many users and huge amount of data
 - Various quality of services for multimedia applications
 - Efficient traffic management



Number of IP service subscribers and number of packets

(Source: MIC in Japan)

2

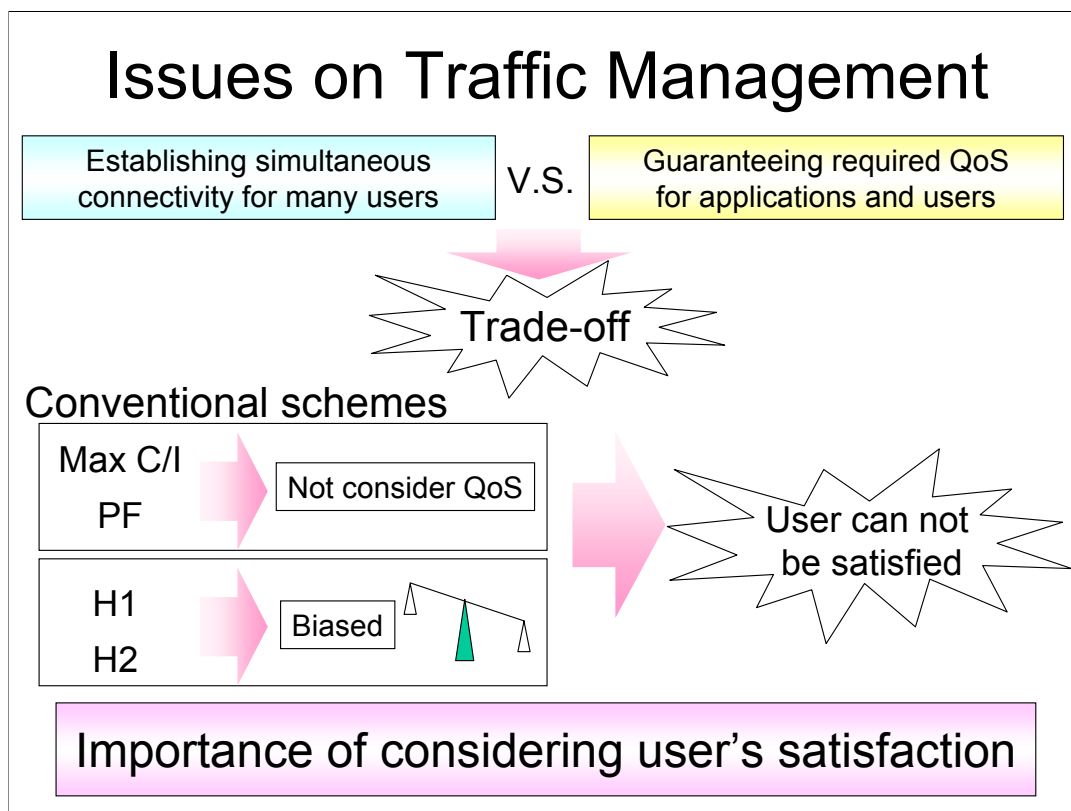
1. Introduction

The next-generation All-IP mobile communication systems require a large capacity as well as guaranteeing various levels of quality of service (QoS) for multimedia applications[1].

According to the recent report published by Ministry of Internal Affairs and Communication (MIC) in Japan, both the number of subscribers to IP service using mobile phones and the number of data packets have been increasing rapidly. To meet such a large amount of traffic demand, wireless systems should provide more capacity. We think there are two perspectives for the requirements of a high-capacity system. One is a perspective from increased data traffic. To cope with this, more base stations should be deployed with narrowing down the area of each cell. Or more specifically, an interference canceller should be installed in order to improve each base station to accommodate more traffic demands. The other is a perspective from the increase of the number of calls in a unit of time. The number of call requests must go up in proportion to the number of subscribers to IP service. Therefore, it is required to make a wireless system available to as many users as possible, for them to be able to connect to it simultaneously. Although the former requirement could be achieved by improving wireless communication technologies, our study focuses on the latter requirement.

On the other hand, the next-generation All-IP mobile communication system is expected to provide various multimedia applications. Such a system has to guarantee a QoS required by users or applications by controlling the data traffic appropriately when providing various application services with a shared channel in a down link to transmit the data packets. In addition, even for the same application service, each user makes a different request. For example, one emphasizes a service quality or another expects a cheaper communication charge. Consequently, it is essential to consider QoS requirements with respect to each user.

Thus, the next-generation All-IP mobile communication environment requires efficient traffic management schemes to achieve both establishing simultaneous connectivity for many users and guaranteeing QoS requirements of various application services.



2. Issues on Traffic Management

Wireless systems, in general, can not accept all users' requests for guaranteed quality because of limited wireless resources. Therefore, there is a trade-off between simultaneous connectivity of users and guarantee of QoS requirements. For instance, a wireless resource allocated to each user might be limited, when many users are establishing connections. On the other hand, some specific users who emphasize quality may occupy most of the wireless resources since it is necessary to reserve sufficient wireless resources for providing the required qualities. This paper proposes a traffic management scheme that performs an appropriate admission control and a rate control in consideration of the above trade-off.

Traffic management is one of the most essential issues in a wireless system and has been studied extensively. In [2], a maximum CIR (Max C/I) scheme is proposed as rate allocation scheme. The Max C/I scheme can maximize a rate of data throughput in a base station by giving a prior transmission to a mobile terminal with a wireless link of a better condition. However, this scheme can not guarantee the required quality by a mobile terminal. Another rate allocation scheme proposed a proportional fairness (PF) scheme [3]. The PF scheme tries to give equal opportunities to transmit packets of mobile terminals based on the calculated criteria such as (instantaneous wireless link quality) / (mean wireless link quality). However, the PF scheme can not guarantee the required quality from mobile terminals, either.

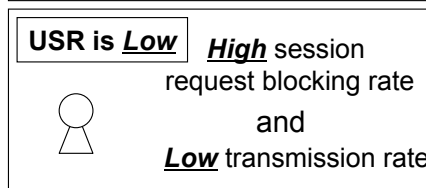
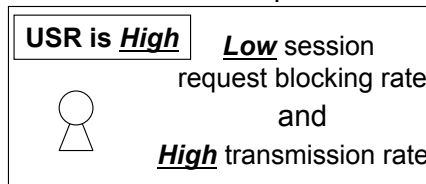
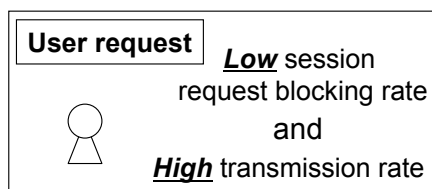
As for a traffic management scheme regarding a trade-off issue mentioned above, H1 and H2 are presented in [4]. H1 can meet the quality requirements of only accepted users, but it might tend to reject many other users' requests. Therefore, H1 accepts a new session request only when the required maximum transmission rate exists in a base station. On the other hand, H2 accepts a new session request only when the required minimum transmission rate exists in a base station. H2 achieves the simultaneous connectivity of many users. However, H2 may not allocate the transmission rate required by users, then it is uncertain whether users are satisfied with the quality of the provided service.

It is of extreme importance for a service provider to provide a service meeting with user's satisfaction in a competitive mobile telecommunication environment. For example, an appropriate solution may exist between H1 and H2 from this standpoint. Then, we propose a new traffic management scheme that can strike the right balance of the trade-off, introducing a concept of user's satisfaction rate.

User's Satisfaction Rate

- What is User's Satisfaction Rate (USR)?
 - a criterion for judging whether a user is satisfied with a provided QoS

For examples



Our challenges

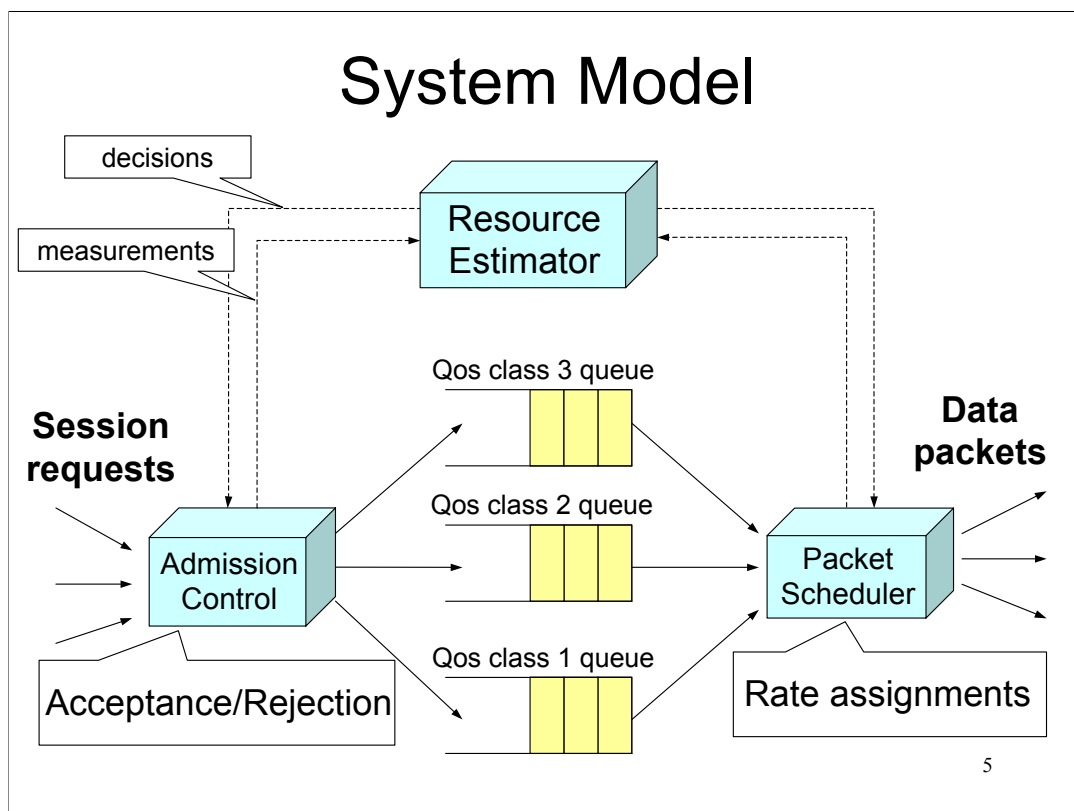
- Improving the minimum value of the USR
- Balancing two sides of the trade-off

3. User's Satisfaction Rate

A user's satisfaction rate (USR) is a criterion for judging whether a user is satisfied with a provided QoS [5]. A high USR means that a service provider can meet the user's requirement. By using the USR, traffic management can make it possible to allocate wireless resources appropriately from users' points of view, in the various communication environments.

As for traffic management schemes using the USR, some studies have been already executed. A fair bandwidth allocation scheme in a wired LAN environment is proposed in [6]. This scheme, however, assumes that the total bandwidth is fixed, and it is not suitable for a wireless communication environment where the available bandwidth fluctuates in a communication link. Meanwhile, there are two traffic management schemes using the USR in a wireless communication environment with a circuit switching. A resource allocation scheme suitable for some specific applications in [7] uses a USR defined in consideration of the response time and the session request blocking rate for the characteristics of some applications. The other channel allocation scheme in [8] uses a USR based on the user's preferences for each application. But the All-IP mobile environment requires guarantee of quality according to the diverse traffic characteristics of a variety of applications unlike the case of the circuit switching.

From the perspective of users, establishing simultaneous connectivity for users and guaranteeing the required QoS are interpreted as a low session request blocking rate and an acquisition of a high transmission rate, respectively. Then, we call the requirement of the session request blocking rate "connectivity" and the requirement of the acquired transmission rate "bandwidth". We assume that a user requests the connectivity and the bandwidth in each session. Moreover, the connectivity and the bandwidth that users request are divided into some classes respectively, since it is difficult to treat a wide variety of users' requests individually. According to the situation of arrival rates of session requests and available resources in a base station, our proposed scheme performs a traffic management based on a USR for each connectivity class and each bandwidth class.



4. Traffic Management Scheme

4.1 System Model

As shown in the above figure, an assumed traffic management system is considered to consist of a resource estimator (RE) to control the system resources to allocate for users, an admission controller (AC) to assess an acceptance or a rejection for a session request, and a rate controller (RC) to allocate an available transmission rate to a packet.

The dotted arrows present the flow of control information. AC and RC regularly advise RE of the measurement values, as represented by the dotted arrows toward RE. AC estimates an arrival rate of the session requests and a mean connection period of each connectivity class. RC measures an available transmission rate in a base station. On the basis of USR, RE decides both of the allowed number of users to connect to the system simultaneously for each connectivity class and a percentage of the available transmission rate for each bandwidth class. RE also informs the determined number of users and the percentage of the available transmission rate to AC and RC respectively, as represented by the dotted arrows toward AC and RC.

On the other hand, the solid arrows present the flow of user information. According to the determined number of the users for each connectivity class, AC assesses whether to accept or reject an arrived session request. RC sends the packets in their proper sequence according to the calculated percentage of the available transmission rate for each bandwidth class.

Definition of USR

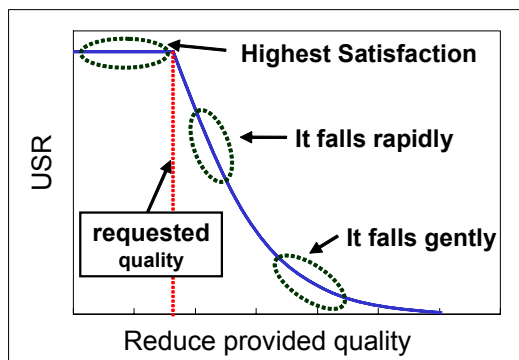


Fig. 1 Proper figure

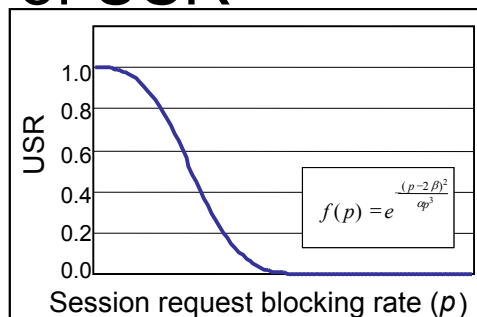


Fig. 2 User's satisfaction rate of "Connectivity"

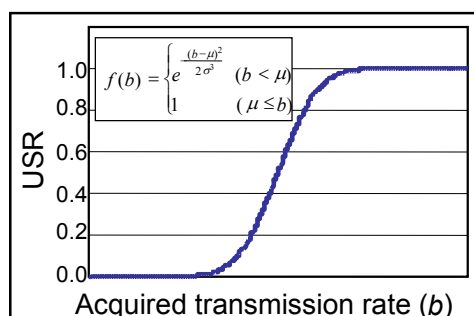


Fig. 3 User's satisfaction rate of "Bandwidth"

4.2 Definition of USR

This section describes two types of USR functions regarding connectivity and bandwidth. Fig.1 shows a general USR function that can be expressed by the variation of a user's satisfaction to a given service. A USR function regarding connectivity is defined as C-USR derived from the session request blocking rate. Another USR function regarding bandwidth is defined as B-USR derived from the acquired transmission rate. The formulation of the two types of USR functions are expressed as follows.

(1) C-USR function: The C-USR function is formulated by the variation of a user's satisfaction to a session request blocking rate, where we assume that C-USR depends on whether a user's session request is accepted or denied. However, from the viewpoint of the standardization, 3GPP (3rd generation partnership project) has not yet specifically established the recommendation of a session request blocking rate in All-IP mobile telecommunications. Then, this paper supposes that C-USR marks 100%, if the session request blocking rate is less than 0.03. We also have conducted a survey in the form of a questionnaire to pick up the characteristics of the C-USR function. The survey adopted a 5-point scale for the session blocking rate posed to the respondents. Fig.2 shows the determined C-USR function based on the results of the questionnaire.

(2) B-USR function: A type of user's request for bandwidth depends on the characteristics of a provided application service. For instance, a user may require a stable bandwidth for the on-demand streaming service. On the other hand, a user may require a quick response time for the web-browsing. The 3GPP has already defined four different types of QoS classes such as conversation class, streaming class, interactive class and background class [9]. In this paper, we formulate a B-USR function specific to the streaming class, since the streaming service will be most prevailing in a future mobile telecommunication market. Although the quality of the streaming service depends on an encoding rate, buffer shortage at a receiver and packet overflow in a network queue, there has been a proposed encoding rate control considering both of the buffer shortage and the packet overflow on a multimedia streaming server [10]. Therefore, the B-USR function regarding the acquired transmission rate in Fig.3 is defined on the basis of the report of quality evaluations about an encoding rate [11], assuming an environment employing such an encoding rate control.

Classification of Users' Requests

- Extraction method: AHP (Analytic Hierarchy Process)
 - A method to express a user's decision numerically

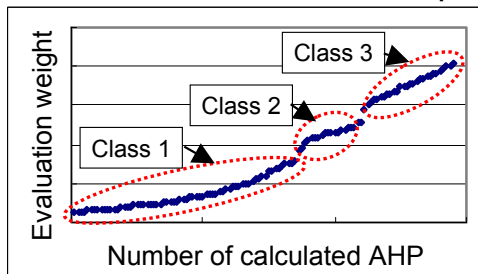
Table 1. An example for calculating the AHP

	Cost	Connectivity	Bandwidth	Evaluation weight
Cost	1	1/5	1/7	0.067
Connectivity	5	1	1/5	0.219
Bandwidth	7	5	1	0.715

The paired comparison table base on an importance level that user inputs

The evaluation weight of each item calculated from paired comparison

- Classification of users' requests



Evaluation weight regarding "Connectivity" is the same as Fig. 1

3 Classes regarding "Bandwidth Class" and "Connectivity Class" respectively

Fig. 1 Example of evaluation weights of "Bandwidth"

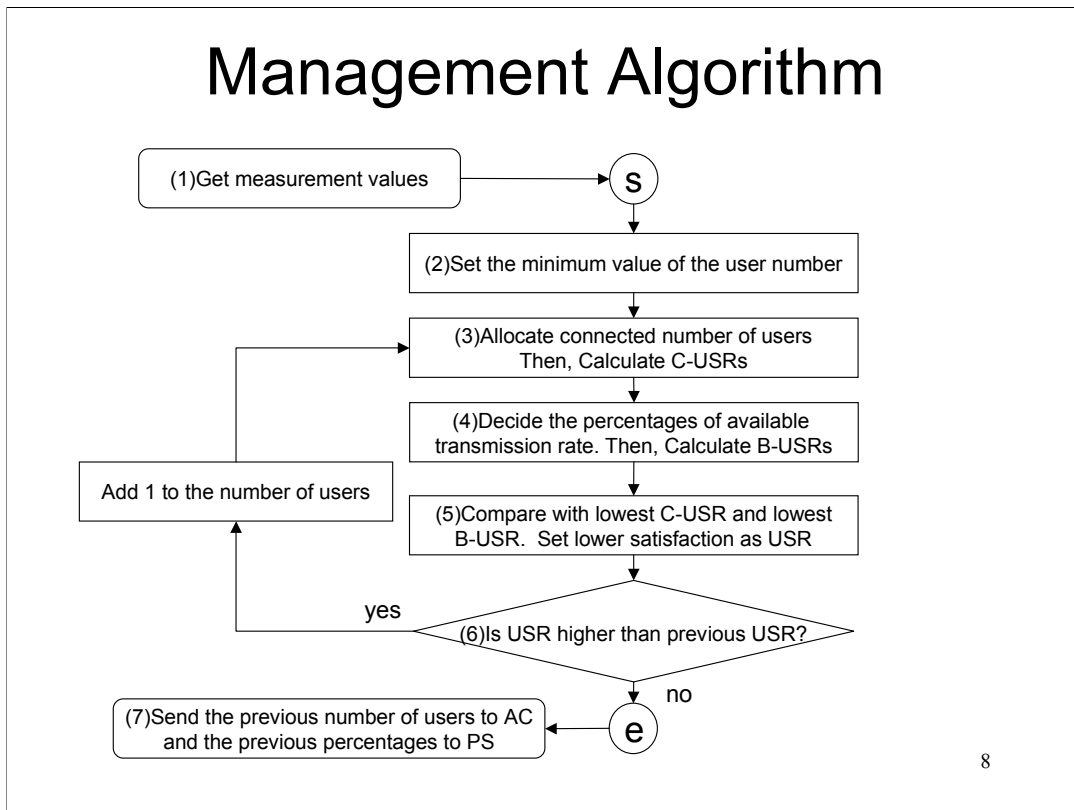
4.3 Classification of USR

The connectivity and the bandwidth that users request are divided into some classes respectively, since it is difficult to treat a wide variety of users' requests individually. The explanation below is the classification procedure based on users' requests.

(1) Extraction of user's request: To classify users' requests, the connectivity and the bandwidth are extracted by employing the Analytic Hierarchy Process (AHP) [12]. The AHP is appropriate for treating our presented trade-off, because AHP quantifies the relative degree of importance regarding plural comparison items as an "evaluation weight". We use "connectivity", "bandwidth" and "cost" as the comparison items. Employing "cost" as a comparison item makes it possible to consider an additional user's request for communication expenses. Table 1 shows an example of the parameters by AHP. At first, a user inputs a value of an importance level restricted in the 5-point scale (i.e. 1, 3, 5, 7 and 9), performing the paired comparison of a row item to a column item. For instance, if a user inputs the importance level of 5 as a result of the paired comparison of "connectivity" in the second row to "cost" in the first column, which means that the connectivity is just more important than the cost for the user. Then, AHP calculates each relative evaluation weight for the row item with the values of the importance levels. The values of the evaluation weight in Table 1 indicate that the user has an more overwhelming need for bandwidth (0.715) than other items.

(2) Classification of users' requests: We classify users' requests on the basis of the evaluation weight of each item calculated by AHP. In this paper, we omit the USR regarding the cost for the simplicity, even if the class of cost can be obtained from the class of connectivity and the class of bandwidth since the total sum of three evaluation weights must be equal to 1. To classify users' requests into some classes, we experimentally calculated all combinations of the importance level by employing the AHP. Fig. 1 shows that the evaluation weights in regard to the bandwidth are plotted in ascending order. The evaluation weights are approximately dispersed into three groups. On the other hand, we similarly classified the connectivity into three classes since the distribution of the evaluation weights regarding the connectivity denotes the same tendency of the one regarding the bandwidth. According to the classification of users' requests, we formulated the USR functions of each connectivity class and each bandwidth class. The higher the class is, the more rapidly the USR decreases.

Management Algorithm

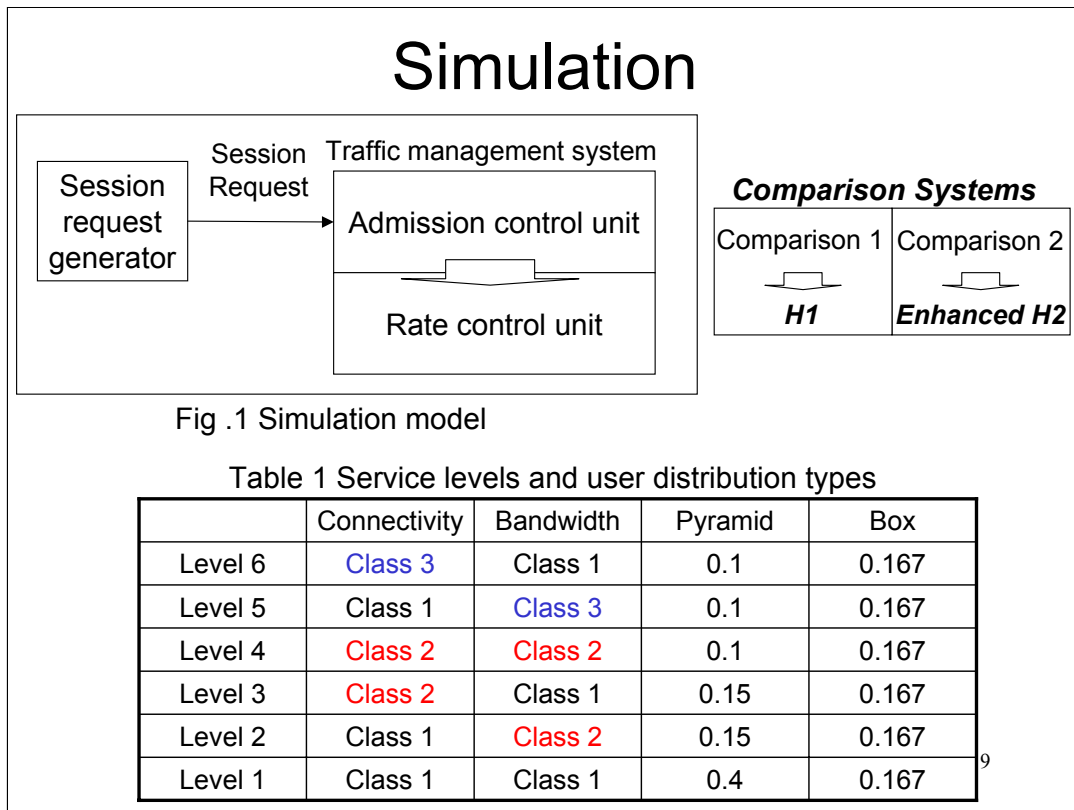


8

4.4 Management Algorithm

This section explains the proposed scheme considering the user's satisfaction that improves the smaller value of C-USR and B-USR. This is because we think that users will not be satisfied when one of the USR values is low, even if the other USR value is high. The proposed scheme appropriately finds the number of simultaneous users connecting to the system by balancing between C-USR and B-USR. For instance, when C-USR is lower than B-USR due to the higher session request blocking rate, the proposed scheme increases the number of the simultaneous users. In contrast, when B-USR is lower than C-USR due to the lower allocated transmission rate, the scheme decreases the number of the simultaneous users. The procedure is shown as below.

- (1) RE gets the session request arrival rate and the connection time for each connectivity class from AC and gets the available transmission rate in a base station from RC.
- (2) RE sets an arbitrary smaller value (e.g. 3) into the initial number of the simultaneous users.
- (3) RE calculates C-USR for each connectivity class from the session request blocking rate derived from the arrival rate, the connection time and the number of the allocated users by employing the Erlang-B formula. This step has the condition to keep C-USR for the higher class above that of the lower class.
- (4) RE calculates B-USR for each bandwidth class from the percentage of the allocated transmission rate under the condition to keep B-USR for the higher class above that of the lower class. The percentage of the allocated transmission rate is determined by estimating the number of users for each bandwidth class on the basis of the number of users for each connectivity class.
- (5) RE chooses the smallest values of C-USR and B-USR among the connectivity classes and the bandwidth classes, respectively. Then RE decides the minimum value of USR (minUSR) by comparing the smallest value of C-USR with the smallest value of B-USR.
- (6) If minUSR is higher than the previous minUSR, RE holds the number of simultaneous users allocated to each connectivity class and the percentage of the allocated transmission rate for each bandwidth class. Then RE repeats the process from the step (3) after adding 1 to the number of simultaneous users. Otherwise, if minUSR is lower than the previous minUSR, RE go ahead with the process to the step (7).
- (7) RE sends the previous number of simultaneous users and the previous percentage of the allocated transmission rate to AC and RC respectively.



5. Simulation

5.1 Simulation environment

We simulated to demonstrate the effectiveness of our proposed system. Fig.1 shows the configuration of the simulator. This simulator assumes a traffic management on a base station which use an omni-directional antenna and that a streaming server control the encoding rate corresponding to the transmission rate allocated to a user. The simulated system has a session request generator and a resource management framework, which consists of an Admission Control unit (AC unit) at the session level and a Rate Control unit (RC unit) at the packet level. AC unit judges the acceptance of a session request transmitted from session request generator, and RC unit schedules packets in queues. The proposed system is compared with two conventional systems. That is, one (Comparison 1) of them is H1 that allocates the required maximum rate to every connected user, and the other (Comparison 2) is enhanced H2. Comparison 2 allocates the minimum rate to every session as same as H2, while Comparison 2 is different from H2 because of allocating the surplus of the available transmission rate to upper bandwidth class by giving higher priority. In this simulation, by comparing C-USR with B-USR in the proposed system, we show that proposed system accomplishes to balance connectivity with bandwidth. In addition, by comparing with the USRs of conventional systems mentioned above, we also show that proposed system improves the minimum value of USR.

The simulation environment is shown as follows.

Down link : Shared Channel

User number : 100

Session request generating probability : 0.1 – 1.0 [request/sec] (Poisson distribution)

Available transmission rate : 1.92 – 3.84Mbps

Application : streaming

Maximum rate : 384kbps

Minimum rate : 32kbps

5.2 Users' requests

Users select their request levels from six service levels, which consist of the combinations of evaluation weights that are calculated by using AHP. Then, we prepare two distribution patterns of user requests. We define the one distribution as the pyramid type; the higher the service level is, the smaller the ratio of user is, and the other as box type; the ratios of all service levels are equal. Table 1 shows six service levels and two distributions.

Balance between “Connectivity” and “Bandwidth”

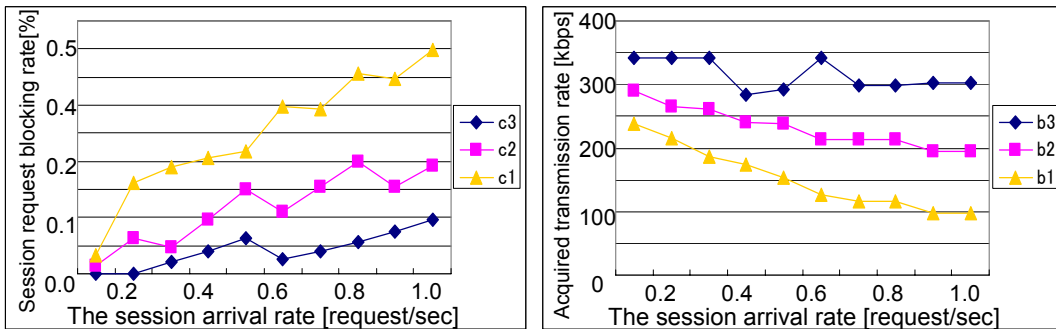


Fig .1 Session request blocking rate of each class Fig .2 Acquired transmission rate of each class

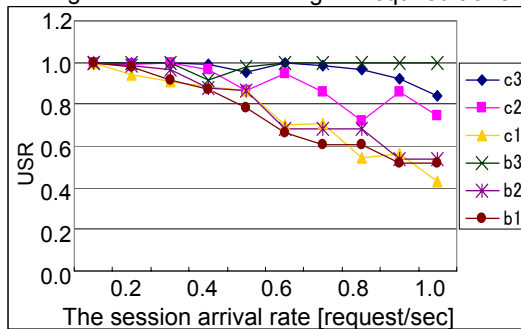


Fig .3 USR regarding connectivity and bandwidth of each class

10

5.3 Balance between “connectivity” and “bandwidth”

We demonstrate the effectiveness of proposed system in terms of the balance between C-USR and B-USR. Above figures show the relations of session request blocking rate for each connectivity class, acquired rate for each bandwidth class and C-USRs and B-USRs to the arrival rates respectively, when the transmission rate of base station is 2.56Mbps.

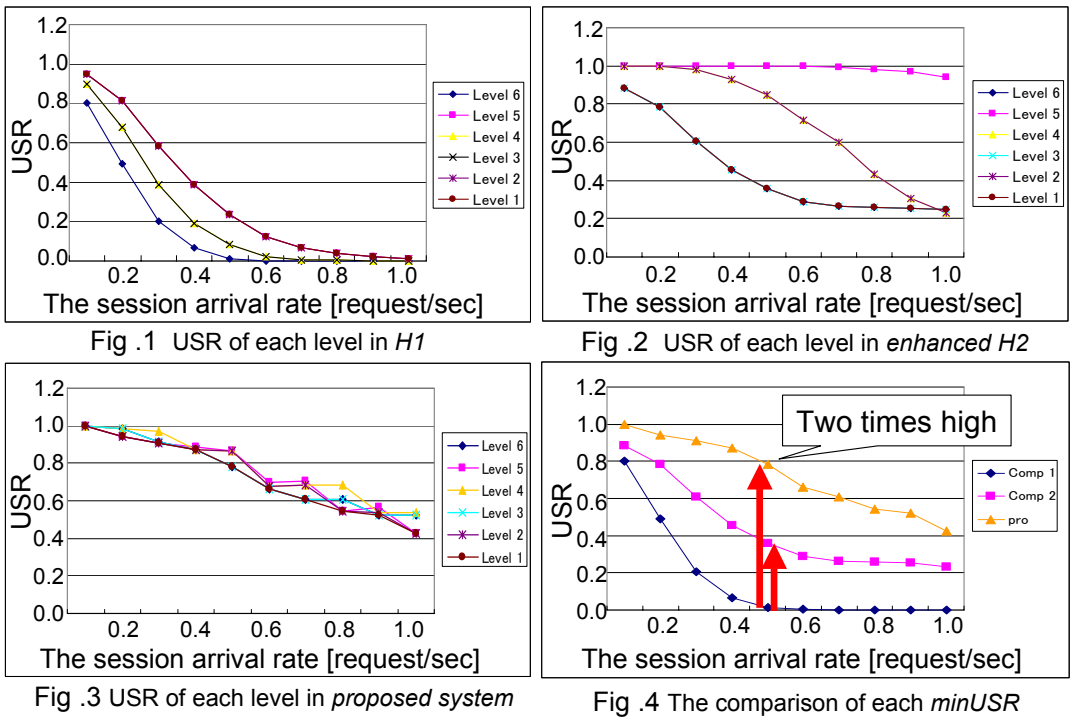
The session request blocking rate of upper connectivity class is lower than that of lower class in Fig.1. The session request blocking rates are ups and downs in places, since the proposed system controls not to excessively allocate the allowed number of simultaneous connection users for the upper class.

Fig.2 shows the transmission rate which the user of each bandwidth class acquires. The higher the bandwidth class is, the higher the acquired transmission rate is. Then, there is no rapid degradation of the acquired transmission rate. Users of bandwidth 3 is provided high transmission rate though there are few degradation of the acquired transmission rate. Users of bandwidth 2 are also kept the acquired transmission rate relatively high, but the acquired transmission rate deteriorates gently. Then, there is no rapid degradation in bandwidth 1 though the acquired transmission rate deteriorates along with the increase of the session arrival rate.

USRs of connectivity classes and bandwidth classes are shown in Fig.3. C-USR of each connectivity class is averagely kept high without providing only a certain class with excessively high or low C-USR. Connectivity 3 constantly obtains high C-USR. C-USR of connectivity 2 comparatively follows C-USR of connectivity 3 though partly going down. Connectivity 1 obtains high C-USR till the arrival rate is 0.5[request/sec]. After that, by toughly minimizing the degradation of C-USR of connectivity 1, there is no rapid degradation, though C-USR of connectivity 1 goes down. B-USRs are averagely kept high without rapid degradation as same as C-USRs.

Thus, the proposed system accomplishes to balance C-USR with B-USR as shown in Fig.3 by not excessively allocating the allowed number of simultaneous connection users or the acquired transmission rate for a certain class and by deciding the total allowed number of simultaneous connection users on the basis of comparison between C-USR and B-USR.

Comparison of The Minimum Value of USRs



5.4 Comparison of the minimum value of USRs

Fig. 1 shows the smaller value of C-USR and B-USR for each service level in Comparison 1. The allowed number of simultaneous connected users is limited, since Comparison 1 allocates the maximum rate that user's requests. Therefore, the smaller value of USR for each service level goes down along with the degradation of USR when the arrival rate increases.

The smaller value of USR for each service level in Comparison 2 is shown in Fig. 2. Comparison 2 can make many users connected by allocating the minimum rate to every session request. The USRs of all levels except the level 5 rapidly deteriorate with the increase of the session arrival rate, though the comparison 2 can get a higher USR of the level 5 since the proposed system assigns the surplus of the transmission rate to a higher bandwidth class by giving a higher priority. Therefore, the smaller values of USR of all levels except the level 5 deteriorate.

Fig. 3 shows the smaller value of USR for each service level in the proposed system. Our proposed system improves the minimum value of USR by comparing with C-USRs and B-USRs. Therefore, the smaller value of USR for each service level deteriorates more gradually than the other systems. In this figure, the smaller values of USR do not go down smoothly.

The comparison with minUSR for each system is shown in Fig. 4. This result demonstrates that the proposed system can keep minUSR twice higher than the comparison 2 when the arrival rate is 0.5 (request/sec).

Thus, this simulation denotes that the proposed system accomplishes to balance between C-USR and B-USR by comparing the session request blocking rate for each connectivity class with the acquired transmission rate for each bandwidth class. The proposed system improves minUSR at least more than two times high in comparison with other conventional systems.

Concluding Remarks

Conclusion

- Proposed a flexible traffic management scheme in consideration of the User's Satisfaction Rate (USR) regarding "connectivity" and "bandwidth".
- Simulation analysis demonstrated that the proposed scheme improves minUSR two times higher than other conventional systems and balances C-USR with B-USR.

Future works

- The traffic managements considering interactive services.
- The admission control scheme taking into consideration the user mobility.

12

6. Concluding remarks

We proposed a novel traffic management scheme considering USRs in regard to establishing simultaneous connectivity for many users and guaranteeing QoS requirements of various application services in the next-generation All-IP mobile environment. Our simulation analysis demonstrated that the proposed scheme improves minUSR two times higher than other conventional systems by balancing C-USR with B-USR. As for the future works, we will discuss some additional issues that can be explored as an extension to this work: (1) Our scheme should also consider the applications in an interactive class defined in 3GPP (e.g. web-browsing and e-commerce). Especially, the scheme has to cope with an application which requires a shorter downloading time. (2) The admission control scheme should be further improved by taking user mobility or a handoff request into consideration. Our scheme will be able to handle the handoff requests as another kind of the admission request by enhancing one.

References

- [1] S. Ohmori, Y. Yamao, and N. Nakajima, "The future generations of mobile communications based on bandwidth access technologies," IEEE Communications Magazine, pp.134-142, Dec. 2000.
- [2] 3GPP, "Physical Layer Aspects of UTRA High Speed Downlink Packet Access," 3G TR 25.848, v4.0.0, Mar. 2001.
- [3] A. Jalali, R. Padovani, R. Pankaj, "Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system," Proc. of IEEE VTC 2000-Spring, Tokyo, vol.3, pp.1854-1858, 2000.
- [4] J.-Y. Jeng, Y.-B. Lin, and W. Chen, "Dynamic scheduling for GSM data services," IEICE Trans. Commun., vol.E80-B, no.2, pp.296-300, Feb. 1997.
- [5] K. mase, "Multimedia Networks and Communications Quality," IEICE, pp.24-28, 1998.
- [6] Y. Hidaka and Y. Teshigawara, "Multimedia Traffic Management Methods in Consideration of End-user Satisfaction," APNOMS1997, Technical Proceedings, pp.275-286, 1997.
- [7] T. Wakuda and Y. Teshigawara, "Traffic Management Methods for SLA/SLM in Consideration of User Satisfactory Levels in the Mobile Environment," APNOMS2000, Technical Proceedings, pp.275-286, 2000.
- [8] T. Katayama and Y. Teshigawara, "A Dynamic Channel Allocating Method Adapted to Users' Preferences in Circuit Switching of W-CDMA," APNOMS2003, Technical Proceedings, pp.477-488, 2003.
- [9] 3GPP, "Quality of Service (QoS) concept and architecture," 3G TS 23.107 v6.1.0, Mar. 2003
- [10] Y. Onoe, H. Hagino, Y. Atsumi, O. Takahashi, H. Komaki, T. Kushida and N. Tamanouchi, "Network Information Based Rate Controls on Multimedia Streaming Servers," Information Processing Society of Japan Journal, vol.44, no.3, pp.625-635, Mar. 2003.
- [11] RealVideo 9, "RealVideo 9 Technical Report," Real Networks, http://service.jp.real.com/help/library/guides/realvideo9/rv9_tech_overview_v1_5_J.pdf, Oct. 2003.
- [12] T. L. Saaty, "The Analytic Hierarchy Process," McGraw-Hill, NewYork, 1980.