Proceedings of APCC2008 copyright (c) 2008 IEICE 08 SB 0083

Efficient Radio-Resource Utilization by User Position Control with Incentive Rewards

Makoto YOSHINO, Ryoichi SHINKUMA, and Tatsuro TAKAHASHI
Communications and Computer Engineering, Graduate School of Informatics, Kyoto University
Yoshidahonmachi, Sakyo-ku, Kyoto, 606-8501 Japan
Email: {shinkuma, ttakahashi}@i.kyoto-u.ac.jp

***Abstract*—Since radio resources such as bandwidth and signal power are limited and shared by users in a service area, all of the users cannot consistently obtain satisfactory radio resources when the number of users in the service area increases past a certain point. A solution for such a problem is user-position control. In the user-position control, the operator informs users of better communication areas (or spots) and navigates them to these positions. However, because of subjective costs caused by subjects moving from their original to a new position, they do not always attempt to move. To motivate users to contribute their resources in network services that require resource contributions for users, incentive-rewarding mechanisms have been proposed. However, there are no mechanisms that distribute rewards appropriately according to subjective factors involving users. Furthermore, since the conventional mechanisms limit how rewards are paid, they are applicable only for the network service they targeted. In this paper, we propose a novel incentive-rewarding mechanism to solve these problems, using an external evaluator and interactive learning agents. We also investigated ways of distributing rewards based on user state and user contributions. We applied the proposed mechanism and reward control to the user-position control, and demonstrated its validity.**

I. Introduction

Since radio resources such as bandwidth and signal power are limited and shared by users in a service area, all of the users cannot consistently obtain satisfactory radio resources when the number of users in the service area increases past a certain point. As a result, if users make new service requests in that area, they could be rejected or their service quality could decrease to unsatisfactory levels. To solve this problem, user-position control has been proposed [1][2].

Figure 1 shows the mechanism for user-position control, where the operator suggests that user *C* move to a position with a better channel. If user *C* obtains a higher transmission rate by moving into a better channel area, as depicted in Fig. 1, the amount of resources consumed will reduce and throughput will be maintained. Then, as illustrated in Fig. 2, the operator can reassign spare resources to new users *D* and *E* or existing users *A*, *B*, and *C*, resulting in greater user satisfaction (utilities). In reassigning resources to existing users, improvements to the user's channel quality will increase the utilities for all existing users. However, this increase in utilities is not guaranteed to compensate for the subjective cost experienced by the user having to move. In reassigning resources to new users, movement to a new area by a user will increase the number of accepted users, resulting in an increase in their utilities. However, the utility of the user who has had to move will not increase. Therefore, in both cases, an incentive mechanism that compensates for this cost and that motivates users to move is necessary.

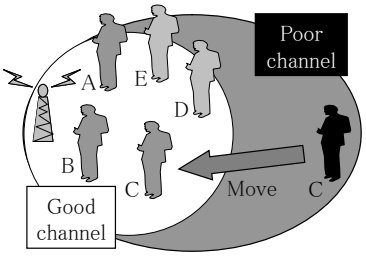

Fig. 1. User-position control.

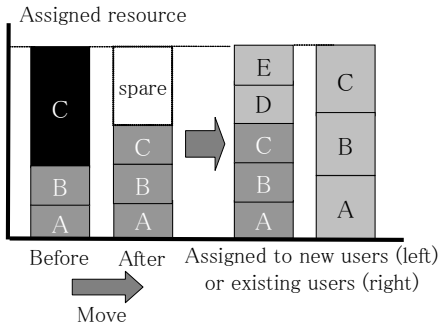

Fig. 2. Resource reassignment by user-position control.

Incentive-rewarding mechanisms for network services have been proposed. They require resource contributions from users such as those in peer-to-peer (P2P) and multi-hop networks; rewarding them with incentives motivates users to contribute to the services [3]-[5]. However, most of these studies have predefined rewards and costs as objective mathematical formulas, and they have discussed issues on a limited definition. This means that no mechanisms distribute rewards to users appropriately according to subjective factors. In addition, rewards can be paid in any form as long as they are equal in value. However, the conventional mechanisms limit how rewards are paid, so they are applicable only for the network service they target.

We propose and describe a novel incentive-rewarding mech-

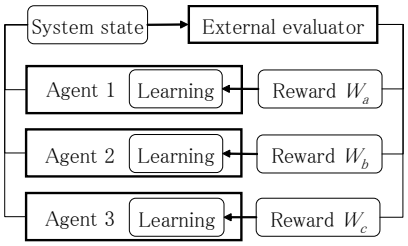

Fig. 3. Multiagent system with external evaluator.

anism that solves these problems using an external evaluator and interactive learning agents. This mechanism is called EMOTIVER (Everyone's MOTivated by incentIVE Reward). We also investigated ways of controlling rewards based on user state and user contributions. We applied EMOTIVER and reward control to user-position control, and a demonstration of its validity is presented.

II. EMOTIVER-Our incentive-rewarding mechanism-

EMOTIVER consists of two parts: an external evaluator and interactive learning agents, both of which are described in detail along with their functions in the following sections.

A. External evaluator

A reward-distribution problem can be considered as a multi-agent system, which consists of multiple learning agents. In a multiagent system, all agents share the rewards obtained from the system state. Each agent selects a behavior (effort level [6]) that maximizes its own expected profit. However, to stabilize the system at a high level of quality, each agent needs to obtain information about the other agents and to choose its own effort level according to that information. Thus, they require complicated algorithms. Bochi et al. proposed an external evaluator to solve this problem [7]. Figure 3 shows that the external evaluator distributes rewards appropriately to agents based on the total rewards obtained from the system state, the behaviors of agents, and the reward-distribution function. Each agent finds the behavior that will maximize its expected profit by reinforcement learning. That is, the role of the external evaluator is to steer the system to achieve high-quality results by using rewards effectively.

B. Interactive learning agent

Even if we only apply the external evaluator stated in the previous section to user-position control, all users are agents and select the effort level that will maximize their expected profit according to the rewards given by the external evaluator. However, searching for the optimal effort level incurs additional costs for users, and they are not always going to find it. To solve this problem, we introduced an interactive learning agent proposed by Lee et al. [8]. This agent first chooses a selectable wireless service and suggests the user to use it. The agent then receives satisfaction/dissatisfaction responses for the service fed back by the user and finally finds the service that maximizes her/his satisfaction. Although Lee et al. used this agent only for a context-aware service in heterogeneous wireless networks, by applying to our incentive mechanism, which is shown in Fig. 4, all users can easily find their optimal effort level through simple interactions with their learning agents.

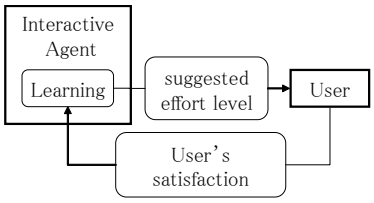

Fig. 4. Interactive learning agent.

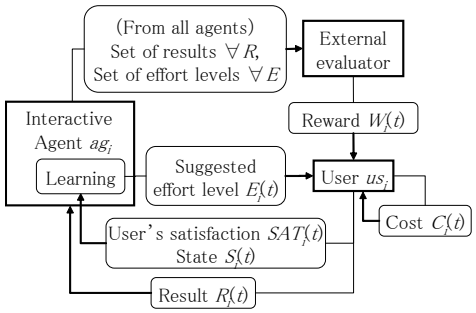

Fig. 5. EMOTIVER.

C. System flowchart

Figure 5 presents a flowchart of EMOTIVER. The flow includes six steps:

1) Interactive agent ag_i chooses the effort level for period t $E_i(t)$ from effort level table T, which lists all selectable effort levels and proposes this to user us_i.
2) us_i behaves according to $E_i(t)$ and produces result $R_i(t)$.
3) ag_i obtains $R_i(t)$ and reports $E_i(t)$ and $R_i(t)$ to the external evaluator.
4) us_i obtains reward $W_i(t)$ from the external evaluator and feeds back her/his satisfaction level $SAT_i(t)$ to ag_i based on $W_i(t)$ and cost $C_i(t)$.
5) ag_i inputs the set of $E_i(t)$, $SAT_i(t)$, and user state $S_i(t)$ in the learning algorithm and proposes the next effort level $E_i(t+1)$ to us_i according to the next user state $S_i(t+1)$.
6) The flow from step 2) to 5) is repeated until the optimal effort level has been found, which us_i can recognize when ag_i does not change the proposed effort level during a certain long period. After that, us_i continues to behave based on his/her optimal effort level.

Satisfaction level $SAT_i(t)$ of us_i is given as:

$$SAT_i(t) = W_i(t) - C_i(t) \tag{1}$$

where $W_i(t)$ is the reward us_i obtained, and $C_i(t)$ is the cost s/he incurred. This relationship is reliable because rewards and costs can be converted into cash and because it has been proved through a demonstration under limited conditions [2]. However, rewards can be paid in any form as long as they are equal in value, i.e., cash, virtual currency, bandwidth, etc. Thus, unlike conventional mechanisms, no predefinition of cost and utility functions is required in EMOTIVER; EMOTIVER supports subjective user factors such as costs and types of satisfaction.

D. Reward distribution

The external evaluator manages the effort levels and the results for all users. On the basis of these results, it distributes rewards to users at the end of period t as follows.

1) The evaluated value is determined from the set of $E_i(t)$ and $R_i(t)$:

$$Eval_i = f(E_i(t), R_i(t)). \tag{2}$$

2) Based on the evaluated values of all users, the reward for us_i is determined.

$$W_i(t) = g(Eval_1, Eval_2, \cdots, Eval_i, \cdots) \tag{3}$$

where the total of rewards $W_{total}(t)$ must be:

$$W_{total}(t) = \sum_i W_i(t). \tag{4}$$

We can control users' effort levels and the system quality by controlling these evaluation functions and the reward-distribution function. In this paper, we used a function proportional to $R_i(t)$ [3]. That is, Eq.(2) and Eq.(3) can be respectively represented as

$$\begin{cases} Eval_i = R_i(t) \\ W_i(t) = W_{total} \times \frac{Eval_i}{\sum_i Eval_i}. \end{cases} \tag{5}$$

III. Simulation

A. Simulation model

In the following subsections, we discuss the application of EMOTIVER and our method of regulating rewards to user-position control. We adopted a simple and versatile multiagent simulation model because, in the future, we want to apply our model to user-position controls in various wireless access systems. Our method cannot easily be compared with the conventional methods mentioned in Sect. I because, as explained in the section, they do not have a mechanism that monitors the satisfactions of users and that distributes rewards according to subjective user factors. Therefore, we discuss the optimality of our reward-distribution method.

1) User-position control: We modeled reassignment of resources to new users using user-position control based on the quality of radio channels explained in Sect. I [2], and Fig. 2. The simulation parameters are listed in Table I. The wireless service areas are separated in 6-Mbps, 12-Mbps and 24-Mbps areas, which correspond to poor, better, and best qualities of channel. We simply calculated throughput from (throughput)=(ratio of bandwidth utilization time)×(transmission rate). That is, for example, a user in a 6-Mbps area requires twice as long the bandwidth utilization time as a user in a 12-Mbps area to obtain the same throughput. Figure 6 shows an example of the user-position control we used in the simulation. When user *2* moves from the 12-Mbps area to the 24-Mbps area and user *3* moves from the 6-Mbps area to the 12-Mbps area, we can reassign a spare bandwidth utilization time squeezed by the movement to users *4* and *5*, whose services have not been accepted yet. Note that, to supress the increase of movement cost, we limit our position control only to the movement to the next better channel-area; users do not directly move from the 6-Mbps area to the 24-Mbps area.

TABLE I
Simulation parameters.

No. of users	50
Transmission rate	6/12/24 Mbps
Service application	1 Mbps

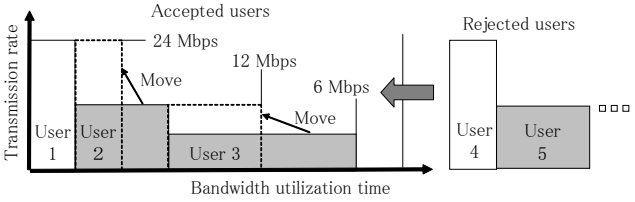

Fig. 6. Model of user-position control (reassignment to new users).

Figure 7 shows the timescale relationships between events, incentive rewarding, and learning. One event is defined in the following four steps.

1) The transmission rate and the service application for each user are randomly determined. The system accepts all users as long as the total bit rate does not exceed the total available bandwidth.
2) The system asks all accepted users to move to the better channel area, and each user responds to or ignores the request according to the effort level determined by her/his agent.
3) After the users have moved, step 2 is applied to new accepted users. However, the system does not ask already moved users to move.
4) When there are no users who can move, the event is completed.

Here, we define the effort level $E_i(t)$ as 'the response rate for requests to move during one set (period t),' which consists of three levels, 0, 0.5, and 1. When a user selects a response rate of 0.5, s/he responds an average of once to two requests to move. We then define the result $R_i(t)$ as 'the number of movements in one set' or as 'an increase in the amount of bandwidth utilization time based on the amount of movement.'

2) Learning algorithm: As outlined in Fig. 7 and stated in the previous section, one set consists of multiple events.

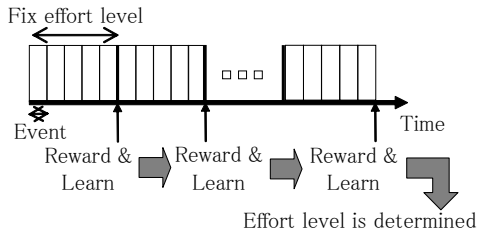

Fig. 7. Timescale relationships between events, incentive rewarding, and learning.

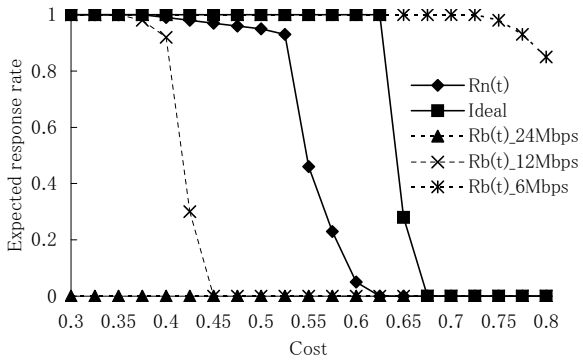

Fig. 8. Expected response rate.

During one set, the response rate is fixed. At every end point of one set, the external evaluator distributes rewards to all users, and all agents learn based on the set of response rate $E_i(t)$, user state $S_i(t)$, and user satisfaction $SAT_i(t)$. The optimal response rate for user us_i has to be determined for each user state; user us_i can have different optimal response rates in each state. In this simulation, whether or not a user can move is considered to be the user state; in the channel areas of 6 or 12 Mbps, the user can move, while in the 24-Mbps area, the user has no better channel area to which s/he can move. Therefore, the Q value can be represented as $Q(E_i(t), S_i(t))$. $SAT_i(t)$ is fed back from users and is used as a reinforcement signal for learning. We used a "profit sharing" updating Q value based on the accumulated value [9]. That is, $Q(E_i(t), S_i(t))$ is updated by $Q(E_i(t), S_i(t)) \leftarrow Q(E_i(t), S_i(t)) + SAT_i(t)$. One set consisted of 50 events, and the simulation time corresponded to 60 sets. We used the ϵ-greedy method ($\epsilon = 0.2$) in the first 20 sets and the greedy method in the rest sets to select the next response rate, where Q value was reset to zero every 20 sets, because users should try all selectable response rates at the beginning and then finally choose the optimal one.

We may reduce the consumed time by improving the learning algorithm or by utilizing the history of user behaviors. A better design for learning algorithms including a reduction in consumed time was not part of the focus of this paper and has been left for future work.

3) Service request model: We assumed multimedia services that would require guaranteed constant bit rates. The total average bit rate required from users was set to 24 Mbps, meaning that all users were accepted when they were in the area with 24 Mbps. In multimedia services that request guaranteed bit rates, the utilities of accepted users can be represented as 1 (=100 %), while the utilities of non-accepted users are 0. Accepted users pay 1 to the system as a willingness to payment (WTP) [10], which is added to the total reward W_{total}.

The cost caused by their moving, on the other hand, can be converted into a willingness to accept compensation (WTA) [10]. That is, we can treat reward and cost on the same scale and represent cost as the relative value of WTP=1. Although the costs experienced by users may differ, we set them uniform so as to analyze the relationship between user contribution and cost.

B. Simulation results

As described in Sect. III-A, the transmission rates and the service applications of users were randomly determined in the simulations. The results and cost for users were influenced by these probabilistic factors, and the relationships are true even in practical environments.

1) Expected response rate: Solid lines in Fig. 8 show the expected response rate for the reward-distribution functions defined in Eq. (5) and Sect. III-A1. The horizontal axis of this figure indicates the cost, while the vertical axis indicates the expected response rate per user. Here, we discuss how the response rate finally chosen by users changes as the cost of movement experienced by users increases or decreases. $R_n(t)$ indicates the expected response rate chosen when $R_i(t)$ is 'the number of movements in one set' and when a user us_i is in the 6 or 12 Mbps area ($S_i(t)$ is 'movable'). When a user us_i is in the 24 Mbps area , all users chose $E_i(t) = 0$ because no area had better channel quality than 24 Mbps ($S_i(t)$ is 'unmovable'). 'Ideal' indicates the maximum expected response rate, which was found in all possible combinations of response rates selected by users.

This figure shows that the reward distribution based on $R_n(t)$ achieved the maximum response rate $E_i(t) = 1$. However, the expected response rate decreased rapidly at $C_i(t) \geq 0.55$ because the number of accepted users increased by about 0.55 times in the simulation due to one user moving, which means the average reward for one user was 0.55. Moreover, compared with 'Ideal', the upper bound of cost where the expected response rate > 0 was smaller.

2) Accepted users: The solid lines in Fig. 9 show the average number of accepted users in the last 20 sets. The horizontal axis of this figure indicates the cost, while the vertical axis indicates the average number of accepted users. Using this metric, we can evaluate how user-position control with incentive rewards can improve system quality.

In $C_i(t) \leq 0.525$, by using $R_n(t)$, we could drastically increase the number of accepted users, which demonstrates the effectiveness of user-position control with incentive rewards. The trends of $R_n(t)$ and 'Ideal' were almost the same as those in Fig. 8.

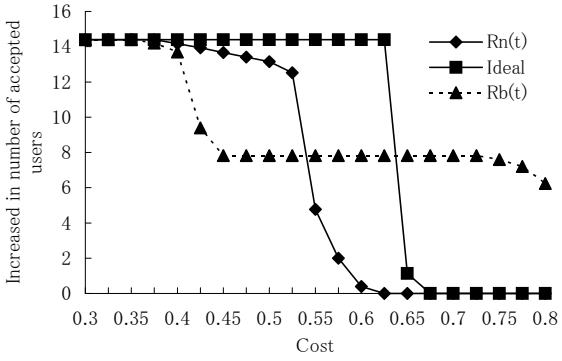

Fig. 9. Increase in number of accepted users

IV. Extension of user state

A. Simulation model

As shown in the previous section, a positive response rate and an increase in the number of accepted users did not occur when the cost was high. This is because users incur cost every time they move, while the total reward increases when the movement increases the amount of spare bandwidth enough to accept a new user. Therefore, we should give more reward to a user when s/he provides a larger amount of spare bandwidth. To achieve it, we introduced another definition of the result 'an increase in the amount of bandwidth utilization time' $R_b(t)$. Moreover, we modified the user state to the channel-area, 6, 12 or 24 Mbps. The reason for this modification is the movement from 6 to 12 Mbps provides a larger amount of spare time than that from 12 to 24 Mbps.

B. Results

1) Expected response rate: The broken lines in Fig. 8 show the expected response rate for the reward-distribution based on $R_b(t)$. $R_b(t)$_6M, for example, indicates the expected response rate chosen when the transmission rate is 6 Mbps. We could not compare this performance with the ideal case because the required calculation time for finding the optimal combination of response rates astronomically increased as the user state was extended. This figure shows that the expected response rate for 6 Mbps was around 1 and was higher than the 'Ideal' regarding movable or unmovable at $C_i(t) \geq 0.575$. Thus, by giving more rewards to users when they provided large spare bandwidth, we can increase the upper-bound cost where the expected effort level is more than 0.

2) Accepted users: The broken line in Fig. 9 shows that the average number of accepted users increased by using $R_b(t)$.

As this figure shows, this reward distribution increased the upper-bound cost when an increase in the number of accepted users occurred. However, in the low-cost region, the increase in the number of accepted users was lower than the 'Ideal'. However, we conclude that $R_b(t)$ is the most suitable for the reward distribution in this evaluation because, as shown in Fig. 9, the performance is not as sensitive to the cost, making it the most attractive characteristic because, in practical environments, the cost experienced by users cannot be easily predicted.

V. Conclusion

We described a novel incentive-rewarding mechanism called 'EMOTIVER,' which uses an external evaluator and interactive agents. These systems take subjective user factors into account and motivate the users to contribute resources to the system. We also investigated ways of controlling rewards. We applied EMOTIVER with its control of rewards to user-position control and tested and validated its effectiveness. On the basis of the simulation results, we concluded that the best way is to distribute rewards based on an increase in the amount of bandwidth and user state. This distribution gives more rewards to users when they provide larger spare resources.

Future work includes building a testbed of a user-position control system and an experimental evaluation of EMOTIVER with people.

Acknowledgement

This work is supported in part by KMRC R&D Grant for Mobile Wireless from Kinki Mobile Radio Center Foundation and the Research Grant from the Okawa Foundation for Information and Telecommunications (no. 07-07).

References

[1] S. Kaneda, et al. "Integrated User and Network Simulation for Traffic Control by Influencing User Behavior," PE-WASUN, pp.99-105, Oct. 2005.

[2] T. Kangawa, et al. "Modeling of Utility and Cost and Its Application for User Position Control based on Radio Channel Quality," IEICE-Transactions on Communications, Vol.J90-B, No.12, pp.1263-1273, Dec. 2007.

[3] W. Tao, et al. "A Novel Incentive Mechanism for P2P Systems," PDCAT, pp.801-803, Dec. 2005.

[4] M. Li, et al. "Pricing to Stimulate Node Cooperation in Wireless Ad Hoc Networks," IEICE-Transactions on Communications, Vol.E90-B, No.7, pp.1640-1650, Jul. 2007.

[5] P. Golle, et al. "Incentives for Sharing in Peer-to-Peer Networks," WELCOM, pp.75-81,Nov. 2001.

[6] P. Milgrom, and J. Roberts, "Economics, Organization and Management," Prentice Hall, 1992.

[7] Y. Bochi, et al. "A direct-indirect reward sharing model in multiagent reinforcement learning," AAMAS, pp.940-941, 2003.

[8] G. Lee, et al. "Learning user preferences for wireless services provisioning," AAMAS, pp.480-487, 2004.

[9] S. Arai, et al. "Multi-agent reinforcement learning for crane control problem:designing rewards for conflict resolution," ISADS, pp.310-317, Mar. 1999.

[10] I.J. Bateman, and K.G. Willis, "Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries," Oxford University Press, 1999.