Design and Implementation of an Optimal Radio Access Network Selection Algorithm Using Mutually Connected Neural Networks

Mikio Hasegawa[1,2], Taichi Takeda[1], and Hiroshi Harada[2]

[1] Tokyo University of Science, Tokyo 102-0073, Japan
[2] National Institute of Information and Communications Technology (NICT), Yokosuka 239-0847, Japan

Abstract. We propose a distributed and autonomous algorithm for radio resource usage optimization in heterogeneous wireless network environment. We introduce optimization dynamics of the mutually connected neural network to optimize average throughput per the terminals and the load balancing among the radio access networks (RANs). The proposed method does not require a server to collect whole information of the network and compute the optimal state of RAN selections for each terminal. We construct a mutually connected neural network by calculating the connection weights and the thresholds of the neural network to autonomously minimize the objective function. By numerical simulations, we show that the proposed algorithm improves both the total and the fairness of the throughput per terminal. Moreover, we implement the proposed algorithm on an experimental wireless network distributively, and verify that the terminals optimize RAN selection autonomously.

1 Introduction

Various radio access networks (RANs) have been developed and deployed. The cellular phone networks provide ubiquitous services and available almost everywhere, but their data communication bit rate is not very high and the cost is relatively expensive. On the other hand, the wireless LAN systems provide high-speed and low cost network access and possible to put access points freely, but it is available only in limited areas, since the coverage of one access point is small. Each RAN has different feature on the connectivity, the transmission speed, the cost per bit, and so on. Therefore, the best RANs for the users to connect to the network always change depending on their situations and available RANs.

Recently, many of the networks are replaced by IP based networks. The cost of the voice over IP communication is much lower than the traditional circuit switched telephone networks. Moreover, the IP enables to exchange various kinds of data, web pages, e-mails, voice, streaming video, etc. By the increase of the demands for the Internet access, the most of RANs provide the Internet connectivity with the global IP access. This means that those RANs are connected to the same core network, the Internet.

J.D. Velásquez et al. (Eds.): KES 2009, Part II, LNAI 5712, pp. 607–614, 2009.
© Springer-Verlag Berlin Heidelberg 2009

Across those RANs connected to the Internet, vertical handover technologies enable to switch the on-going sessions on one of the RAN to other RAN without interruption of the session [1,2]. The mobile IP[3] enables to switch the IP address of a mobile terminal for the session, seamlessly. IEEE 802.21 [4] provides a common interface to the upper layer protocols to control different kinds of RAN interfaces. By using those technologies, it becomes possible to seamlessly handover the sessions among different kinds of RANs.

The best RANs for each user always change depending on the user's location and situation, the network traffic load, the available and required QoS and so on. By vertical handover among different RANs, it becomes possible to optimize the radio resource usage of the whole wireless network environment. The architecture to exchange information required for radio resource usage optimization has been already standardized as IEEE 1900.4 [5]. To find the best RAN for all users to optimize the radio resource usage according to such information becomes a combinatorial optimization problem. There are a lot of researches to improve various factors, such as user throughput, load balancing, user QoS optimization and so on. To improve those factors, some of those researches utilize mutually connected neural network [6] to solve those combinatorial optimization problems [7,8]. Since the mutually connected neural network solves the problem by its autonomous and distributed dynamics, there is no need to run the algorithm at a centralized server with a heavy computational cost, and also no need to collect all of information of large scale network to one centralized server.

In this paper, we evaluate the effectiveness of the neural network approach by computer simulations and real experiments on an experimental wireless network. We apply this optimization approach to load balancing of the traffic, which improves fairness for the users and RANs. For the implementation for real experiment, we use the Cognitive Wireless Cloud (CWC) system [10], and verify the capability of the proposed approach.

2 Load Balancing Based on Neural Network Dynamics

There are a lot of factors to be optimized for the RAN selection in heterogeneous wireless network environment. In IEEE 1900.4 [5], various kinds of information are defined to be exchanged between the network side and the terminal side to choose the best wireless links to be connected. In this paper, we examine the performance and effectiveness of the neural network based algorithm by optimizing the load balancing while keeping maximization of the throughput per user.

First, we define the available throughput per user by an equation. For simplifying the experiments in this paper, we assume that all the terminals are communicating by a best-effort type application, and capacity of each access point is shared equally among the terminals connected to the same access point. Under such an assumption, available throughput for each terminal can be approximately defined as $T_i(t) = C_{h_{\mathrm{link}}(i)}/N^{\mathrm{AP}}_{h_{\mathrm{link}}(i)}(t)$, where $N^{\mathrm{AP}}_j(t)$ is the number of terminals connected to the access point j, C_j is the total of the throughput which the access point j can provide, $h_{\mathrm{link}}(i)$ is the access point which the terminal i is currently connecting.

In order to optimize the fairness and the total of the throughput at the same time, we use the following objective function,

$$E_{\mathrm{OBJ}}(t) = \sum_{i=1}^{N_{\mathrm{m}}} \frac{1}{T_i(t)} = \sum_{i=1}^{N_{\mathrm{m}}} \frac{N^{\mathrm{AP}}_{h_{\mathrm{link}}(i)}(t)}{C_{h_{\mathrm{link}}(i)}}, \tag{1}$$

where N_{m} is the number of mobile terminals in the network. By minimizing this simple function, the fairness and the total of the throughput can be optimized at the same time. Minimization of the reciprocal of the throughput $T_i(t)$ means maximization of the throughput. Moreover, the value of $E_{\mathrm{OBJ}}(t)$ becomes smallest in the case that all $T_i(t)$ becomes equal, when the total $\sum_{i=1}^{N_{\mathrm{m}}} T_i(t)$ is fixed. The problem is formulated as a combinatorial optimization problem that finds an optimal state of RAN selection for each terminal.

In order to minimize this simple function autonomously without collecting whole information and computing everything at one server, we introduce optimization dynamics of the mutually connected neural network. It is well-known that the energy function of the mutually connected neural network,

$$E_{\mathrm{NN}}(t) = -\frac{1}{2} \sum_{i=1}^{N_{\mathrm{m}}} \sum_{j=1}^{N_{\mathrm{AP}}} \sum_{k=1}^{N_{\mathrm{m}}} \sum_{l=1}^{N_{\mathrm{AP}}} W_{ijkl} x_{ij}(t) x_{kl}(t) + \sum_{i=1}^{N_{\mathrm{m}}} \sum_{j=1}^{N_{\mathrm{AP}}} \theta_{ij} x_{ij}(t). \tag{2}$$

always decreased and converges to a state corresponding to a minimum of this energy function by a typical neuronal update , such as the following equation,

$$x_{ij}(t+1) = \begin{cases} 1 & \text{for } \displaystyle\sum_{k=1}^{N_{\mathrm{m}}} \sum_{l=1}^{N_{\mathrm{AP}}} W_{ijkl} x_{kl}(t) > \theta_{ij}, \\ 0 & \text{otherwise}, \end{cases} \tag{3}$$

where, $x_{ij}(t+1)$ is the output of the (i, j) th neuron at time t, W_{ijkl} is the connection weight between the (i, j) th and (k, l) th neurons, θ_{ij} is the threshold of the (i, j) th neuron, respectively. The conditions for this convergence are that the weights of the self feedback connections are 0, $w_{ijij} = 0$, that the weights of the connections between the same pairs of neurons are equal, $w_{ijkl} = w_{klij}$ and that each neurons should be updated asynchronously.

To apply this neural network to solution search in a combinatorial optimization problem, first we have to define the relation between each solution and the firing pattern of the neural network. Since the problem is to find the wireless links which should be selected, we relate the firing of the (i, j) th neuron with an establishment of the wireless link between the terminal i and the access point j as shown in Fig. 1.

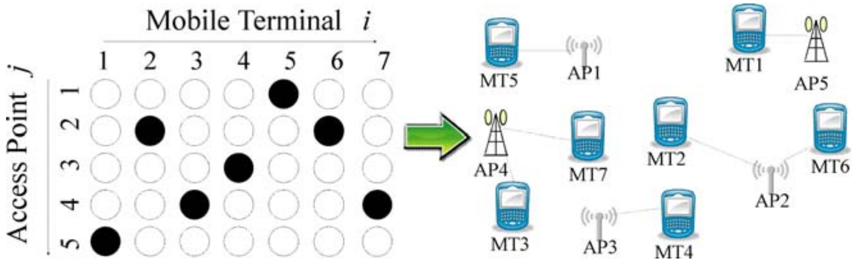

Fig. 1. Relation between firings of the neurons and establishments of the wireless links

Based on the relation described in Fig. 1, Eq. (1) can be transformed to a following form, a function of the state of neurons $x_{ij}(t)$.

$$E_{\mathrm{OBJ}}(t) = \sum_{i=1}^{N_{\mathrm{m}}} \sum_{j=1}^{N_{\mathrm{AP}}} \sum_{k=1}^{N_{\mathrm{m}}} \sum_{j=1}^{N_{\mathrm{AP}}} \frac{1}{C_j}(1-\delta_{ik})\delta_{jl} x_{ij}(t) x_{kl}(t) + \sum_{i=1}^{N_{\mathrm{m}}} \sum_{j=1}^{N_{\mathrm{AP}}} \frac{1}{C_j} x_{ij}(t). \quad (4)$$

By comparing Eqs. (2) and (4), we can obtain the connection weights and threshold to minimize Eq. (1), as follows,

$$W_{ijkl} = -2\frac{1}{C_j}(1-\delta_{ik})\delta_{jl}, \quad (5)$$

$$\theta_{ij} = \frac{1}{C_j}. \quad (6)$$

In the transformation from Eq. (1) to Eq. (4), we need to be careful to avoid self-feedback connection being larger than 0 to satisfy the condition of minimization on the energy function described above.

By autonomously updating each neuron by Eq. (3) with these obtained values, the state of the whole wireless network converges to an optimum state. In order to run this algorithm without centralized computation, we distribute the neurons to each corresponding terminal. Each terminal updates assigned neuron autonomously, makes a handover decision according to the state of the neurons, and hands over to the corresponding selected access point. This decentralized process optimizes the radio resource usage without any centralized computation.

In this paper's experiments, we assume that each terminal can establish only one wireless link with one access point at the same time. To satisfy such constraint, we need to control the number of firings one for each terminal. Usually, in the optimization neural network approach, we introduce a constraint term into the energy function. However, it sometimes could not be satisfied by local minimum problems and fatal infeasible solutions are obtained frequently. Therefore, in this paper, we introduce a maximum firing neuron,

$$x_{ij}(t+1) = \begin{cases} 1 & \text{if } y_{ij}(t+1) = \max\{y_{i1}(t+1), \ldots, y_{iN_{AP}}(t)\}, \\ 0 & \text{otherwise,} \end{cases} \quad (7)$$

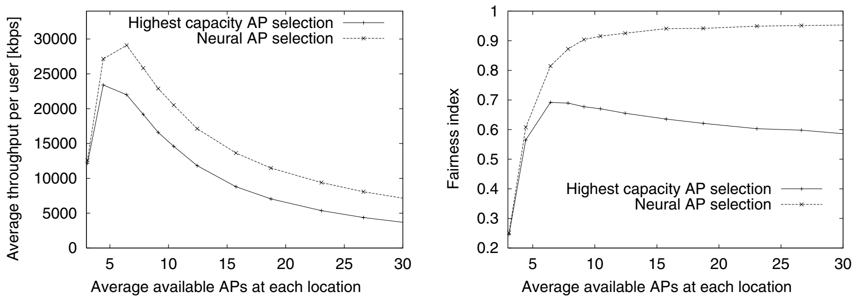

Fig. 2. Average throughput and fairness index of the proposed algorithm based on the neural network dynamics

where, $y_{ij}(t+1) = \sum_{k=1}^{N_{\mathrm{m}}} \sum_{l=1}^{N_{\mathrm{AP}}} W_{ijkl}^{A} x_{kl}(t) - \theta_{ij}^{A},$. By this update equation, we can keep one firing for each terminal.

The performance of the proposed optimization method for throughput maximization and load balancing is evaluated by comparing with the case that each terminal selects an access point which provides highest capacity in its location. To evaluate load balancing performance, we introduce the Jain's fairness index [11],

$$J = \frac{\left(\sum_{i=1}^{N_{\mathrm{m}}} T_i(t) \right)^2}{N_{\mathrm{m}} \sum_{i=1}^{N_{\mathrm{m}}} \left(T_i(t)\right)^2}. \tag{8}$$

The results of the average throughput and the fairness index obtained by computer simulations are shown in Fig. 2. From the Fig. 2, we can confirm that the proposed method, which does not require any centralized computation, is effective for load balancing with improving the throughput, since both the average throughput and the fairness index could be improved. We have tested the proposed algorithm in the case with up to 200000 terminals.

3 Design and Implementation of a Neural Network Based RAN Selection Algorithm

In our implementation of the algorithm, the neurons are updated on each terminal distributively and autonomously. For each terminal, the neurons on the corresponding column in the left figure of Fig.1 are assigned. In real computation, each terminal has to calculate only for a limited number of neurons corresponding to available access points, which are detectable and reachable for the terminal, because we cannot establish a wireless link even if the neurons corresponding to unavailable links fire. By omitting those neurons corresponding to unavailable wireless links, scalability of the proposed system can be much improved.

In this paper, as an experimental wireless network to implement our algorithm, we use the CWC system [9,10]. This experimental system covers any functionality defined in IEEE 1900.4. In this system, various kinds of context information can be exchanged between the Network Reconfiguration Manager (NRM) on the network side and the Terminal Reconfiguration Manager (TRM) on the terminal side, via the Radio Enabler (RE).

In CWC system, three types of wireless network interfaces are defined. One of them is used as a common signaling channel [1,2] (RE in IEEE 1900.4) to exchange various context information. The second one is used to discover the available RANs. The last one is used as the RANs for the data communications. The terminal can seamlessly handover among different RANs, by switching care-of IP addresses for IP-in-IP capsuling packets between the mobility manager and the mobile terminal, that is almost the same procedure as the mobile IP[3]. The CWC system has additional function which enables IP address switching with multi-link aggregation support, but we do not consider it in this paper. We have developed the mobility manager, NRM, and the mobile terminals on the Linux operating system. To the mobile terminals installed on laptop PCs, we can attach various RAN interfaces which provides connectivity to the global Internet with a global IP address, by wireless LAN, 3G cellular systems, PHS and so on. In the following experiments, we use the wireless LANs for evaluating our algorithm based on the neural network dynamics.

In order to update each neuron, each terminal needs to obtain the state of other neurons which have non-zero connection weight with it. The states of those neurons can be derived from the information of each terminal's connecting access point, because the relation between the neurons' states and the selected wireless link is clearly defined so in Fig. 1. Therefore, in the implemented experimental system, each terminal receives such wireless link information of the other terminals, which have the neurons connected to the updating one. The information is transmitted to each mobile terminal's TRM via the NRM. Using such collected information, each terminal updates their neurons. According to the updated states of the neurons, each terminal autonomously selects an access point, and hands over to the selected one.

In the experiment described in the followings, we have used 4 wireless LAN access points and 8 mobile terminals. One of the access points and all of the terminals are placed in a same room. Other three access points are placed in another room. A scenario of this experiment is as follows. The mobile terminals, MT1, MT3, MT5, MT7 and MT8, are initially connected to one of those access points and communicating by the best-effort protocol. Between 50 to 100 seconds after the start of the experiment, the mobile terminals, MT2, MT4 and MT6 start their communications, and the additionally connected to the network. We observe the behavior of the algorithms in this scenario.

For a comparison, we have implemented and tested two algorithms. The first algorithm is that each terminal autonomously selects an access point which has

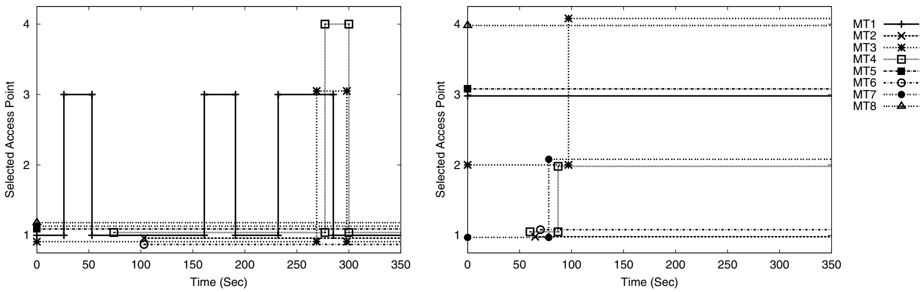

Fig. 3. Time series of selected access points by two algorithms, a generic algorithm in which each mobile terminal selects an access point with the strongest RSSI, on the left, and the proposed algorithm in which each mobile terminal selects an access point based on neural network dynamics optimizing the fairness, on the right

strongest RSSI. The second one is based on the proposed neural network dynamics described in Sec. II. Figure 3 shows the time series of the selected access points of each terminal in those two algorithms. From the Fig. 3 (left), a strongest RSSI selection tends to select the access point 1 which is located in the same room. On the other hand, in the case of proposed neural network based algorithm shown in Fig. 3 (right), selected access point is balanced. After joining of MT2, MT4 and MT6, it took only few neuron updates to converge to an optimal state that the loads of the access points are balanced.

It should be noted that our algorithm does not require any centralized computation to achieve the optimal state. Each terminal exchanges context information corresponding to the neuron state, updates the state of the neurons based on the context information, selects an access point according to the updated state of the neuron, and hands over to the selected the access point. This is distributed and autonomous process. Although we have tested this algorithm in the experimental network with a limited size, we have already shown that this algorithm performs well also in large-scale network by computer simulations.

4 Conclusion

In this paper, we have applied the distributed optimization dynamics of the mutually connected neural network to load balancing of the wireless networks. Since the proposed algorithm does not require any centralized computation, it is suitable for distributed networks, such as the heterogeneous wireless network environment in which each network is managed by different operator. We have shown that the proposed distributed algorithm can optimize fairness of the throughput in a large-scale wireless network, by computer simulations.

Furthermore, we have developed an experimental network to verify the effectiveness of the proposed optimization framework. By comparing our algorithm with a general network selection, we have shown that it is possible to optimize total network resource usage without any centralized decisions or computations.

In this paper, although we have applied our proposed framework only to the load balancing problem, it is applicable to various kinds of optimization problems. In Refs. [12] and [13], we have shown that our distributed algorithm based on the neural network can also optimize other objective functions, such as costs, power consumption and so on, by using more complicated model. In our future work, we are going to apply improved version suitable for more realistic cases in radio resource usage optimization in heterogeneous wireless networks, with evaluations in real experimental wireless networks.

References

1. Wu, G., Havinga, P., Mizuno, M.: MIRAI Architecture for Heterogeneous Networks. IEEE Comm. Mag., 126–134 (2002)
2. Inoue, M., Mahmud, K., Murakami, H., Hasegawa, M., Morikawa, H.: Novel Out-Of-Band Signaling for Seamless Interworking between Heterogeneous Networks. IEEE Wireless Commun. 11, 56–63 (2004)
3. Perkins, C.: IP Mobility Support for IPv4. IETF RFC 3344 (2002); Johnson, D., Perkins, C., Arkko, J.: Mobility Support in IPv6. IETF RFC 3775 (2004)
4. IEEE Std. 802.21: IEEE Standard for Local and metropolitan area networks- Part 21: Media Independent Handover (2009)
5. IEEE Std. 1900.4: IEEE Standard for Architectural Building Blocks Enabling Network-Device Distributed Decision Making for Optimized Radio Resource Usage in Heterogeneous Wireless Access Networks (2009)
6. Hopfield, J.J., Tank, D.W.: Neural Computation of Decisions in Optimization Problems. Biological Cybernetics 52, 141–152 (1985)
7. Gomez-Barquero, D., et al.: Hopfield Neural Network-Based Approach for Joint Radio Resource Allocation in Heterogeneous Wireless Networks. In: Proc. of IEEE Vehicular Technology Conference Fall (2006)
8. Garcia, N., Perez-Romero, J., Agutsi, R.: A new CRRM scheduling algorithm for heterogeneous networks using Hopfield Neural Networks. In: Proc. of WPMC (2006)
9. Harada, H., et al.: A Software Defined Cognitive Radio System: Cognitive Wireless Cloud. In: Proc. of IEEE Globecom (2007)
10. Ishizu, K., et al.: Design and Implementation of Cognitive Wireless Network based on IEEE P1900.4. In: Proc. of SDR workshop (2008)
11. Jain, R., Chiu, D., Hawe, W.: A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer System. DEC Technical Report 301 (1984)
12. Hasegawa, M., et al.: Application of Higher Order Neural Network Dynamics to Distributed Radio Resource Usage Optimization of Cognitive Wireless Networks. In: Köppen, M., Kasabov, N., Coghill, G. (eds.) ICONIP 2008, Part I. LNCS, vol. 5506, pp. 851–858. Springer, Heidelberg (2008)
13. Hasegawa, M., Tran, H., Miyamoto, G., Murata, Y., Harada, H., Kato, S.: User-Centric Optimum Radio Access Selection in Heterogeneous Wireless Networks based on Neural Network Dynamics. In: Proc. of IEEE Wireless Communication and Network Conference (2008)