

논문 2012-49TC-6-5

Q-Learning을 이용한 릴레이 선택 기법 (A Relay Selection Scheme with Q-Learning)

정 홍 규*, 김 광 열*, 신 요 안**

(Hongkyu Jung, Kwangyul Kim, and Yoan Shin)

요 약

차세대 무선통신 시스템에서 다중 경로 페이딩의 영향을 효율적으로 감소시키기 위한 방법으로 최근 협력통신 시스템이 각광을 받고 있다. 협력통신 시스템은 정보를 전송하기 위해서 다양한 페이딩 계수를 가지고 있는 협력 릴레이를 사용하기 때문에, 모든 릴레이를 협력통신에 참여 시키는 것은 자원의 낭비를 초래한다. 그러므로 무선자원을 효율적으로 사용하기 위해서는 최적의 릴레이를 선택적으로 사용할 필요가 있다. 본 논문에서는 무선 협력통신 네트워크에서 발생하는 이러한 문제를 해결하기 위하여 Q-Learning 알고리즘을 이용한 협력 릴레이 선택 기법을 제안한다. Q-Learning에서는 자가 학습을 위해서 상태, 행동, 그리고 보상에 대한 파라미터를 정의한다. 이러한 파라미터가 잘 정의 될 때 Q-Learning을 이용하여 우수한 통신 성능을 얻을 수 있다. Q-Learning 알고리즘의 우수성을 보이기 위해서, 수학적 분석을 통해서 최적의 협력 릴레이를 얻는 기법과 통신 성능을 비교하였다. 모의실험 결과, 제안된 기법에서 Q-Learning 알고리즘 내의 보상을 주는 방식에 따라, 비교 기법과 유사한 심벌오율 성능을 얻으면서 보다 더 적은 협력 릴레이를 선택하는 것을 보였다. 따라서 본 논문에서 제안된 기법은 다수의 릴레이를 사용하는 차세대 무선통신 시스템의 성능 향상을 위한 좋은 접근 방식의 하나로 판단된다.

Abstract

As a scheme to efficiently reduce the effects of multipath fading in next generation wireless communication systems, cooperative communication systems have recently come into the spotlight. Since these cooperative communication systems use cooperative relays with diverse fading coefficients to transmit information, having all relays participate in cooperative communication may result in unnecessary waste of resources, and thus relay selection schemes are required to efficiently use wireless resources. In this paper, we propose an efficient relay selection scheme through self-learning in cooperative wireless networks using Q-learning algorithm. In this scheme, we define states, actions and two rewards to achieve good SER (Symbol Error Rate) performance, while selecting a small number of cooperative relays. When these parameters are well-defined, we can obtain good performance. For demonstrating the superiority of the proposed Q-learning, We compared the proposed scheme with Q-learning and a relay selection scheme with a mathematical analysis. The simulation results show that, compared to a scheme that obtains optimum relays through a mathematical analysis, the proposed scheme uses resources efficiently by using smaller numbers of relays with comparable SER performance. According to these simulation results, the proposed scheme can be considered as a good attempt for future wireless communication.

Keywords: Cooperative Communication, Relay Selection, Self-Learning, Q-Learning, Amplify and Forward

* 학생회원, ** 평생회원, 숭실대학교 정보통신전자공학부
(School of Electronic Engineering, Soongsil University)

※ 본 연구는 지식경제부 및 정보통신산업진흥원의 대학IT연구센터지원사업 (NIPA-2012-H0301-12- 1005) 및 교육과학기술부의 재원으로 한국연구재단의 중견연구자지원사업의 지원을 받아 수행된 연구임 (No. 2012-17220103).
접수일자: 2012년 4월 28일, 수정완료일: 2012년 6월 14일

I. 서 론

무선통신 시스템에서 공간 다이버시티 이득을 얻기 위한 방법으로 다중 릴레이를 이용한 협력통신이 최근에 많은 주목을 받고 있다. 협력 릴레이를 이용하는 방법은 협력통신을 구현하기 위한 중요한 접근 방법 가운데 하나이다. 지금까지 협력 릴레이를 이용하기 위해서 다양한 프로토콜이 제안되어왔다^[1]. 그 중에서도, 릴레이가 소스로부터 전달 받은 신호를 단순히 증폭시켜서 목적지로 재전송하는 AF (Amplify-and-Forward) 방식과 소스로부터 전달 받은 신호를 복호 후에 전송하는 DF (Decode-and-Forward) 방식이 협력통신을 위해서 주로 사용된다.

협력통신 네트워크의 이득을 최대화시키기 위한 핵심은 무선자원을 효율적으로 이용하는 것이다. 특히, 협력통신 네트워크에서 통신 성능을 개선시키고 전력과 대역폭 자원을 효율적으로 사용하기 위해서 전력 할당과 릴레이 선택에 관한 문제를 해결하기 위한 연구가 활발하게 진행되고 있다. 지금까지 많은 논문들이 다중 릴레이로 구성된 협력통신 네트워크에서 수학적 분석으로 최적화 문제를 해결하는 것에 의해 협력 릴레이를 선택하고 최적의 전력을 할당하는 기법들을 연구해왔다^[2-6].

하지만 이러한 수학적 분석은 결과적으로 이상적인 최적의 심벌오율 (Symbol Error Rate; SER) 성능을 구할 수 있으나, 실제 릴레이 선택을 위해서는 대단히 높은 복잡도를 가지는 연산을 수행해야 하는 단점이 있다. 또한 수학적 분석을 통해서 결정된 협력 릴레이 개수 보다 더 적은 릴레이를 사용하면서 최적의 SER 성능과 매우 유사한 성능을 보이는 통신 시스템을 구축할 수 있음에도 불구하고, 수학적 분석을 통한 릴레이 선택 기법으로 구축된 협력 시스템은 이러한 상황에 적응적으로 동작하지 못한다.

따라서 본 논문에서는 이러한 문제를 해결하기 위한 방안으로 Q-Learning을 이용한 자가 학습 릴레이 선택 기법을 제안하였다. Q-Learning은 기계 학습 분야에서 강화 학습 (Reinforcement Learning)에 속하는 알고리즘 중에 하나로^[7], 최근에 또 하나의 차세대 무선통신 시스템으로 주목 받고 있는 인지 무선통신 (Cognitive Radio)^[8-9]에 적용하는 연구가 많이 진행되고 있다. 특히, 다중 사용자와 다중 채널 인지 무선통신 시스템에

서 이차 사용자 사이의 협상 없이 채널을 선택하기 위해서 Q-Learning이 사용 되었고^[10], 인지 무선통신에서 사용자들의 스펙트럼 센싱 시간을 공유하고 전력을 효율적으로 할당하기 위해서 분산된 Q-Learning 알고리즘이 사용되었다^[11]. 한편, [12]에서는 인지 무선통신 기술을 사용하는 멀티미디어 센서 네트워크에서 협력 릴레이를 선택하기 위하여 Q-Learning이 사용되었다. 하지만 [12]은 Q-Learning을 위한 보상값이 데이터 비트를 받고 보내는데 필요한 에너지로 구성되는 등 주로 MAC 계층 관점에서 연구가 수행되었고, 협력 릴레이를 선택하기 위해서 Q-Learning을 이용하는 전반적인 전략은 잘 구성되어 있으나 상태와 행동 등이 명확하게 정의가 되어있지 않다. 반면에 본 논문에서는 물리 계층 관점에서 Q-Learning을 위한 상태, 행동, 보상을 명확하게 정의하고 전체 기법을 제안하였다는데 차별성이 있다. 본 논문의 제안 기법에 대한 모의실험 결과를 통해, Q-Learning을 이용한 제안 기법은 수학적 분석을 통해서 릴레이를 선택하는 기법과 비교했을 때 시스템의 복잡도를 효율적으로 저감시키며 유사한 SER 성능을 보였다. 특히, 이러한 결과는 Q-Learning 알고리즘이 차세대 무선통신 시스템의 성능 향상을 위해서 효과적으로 사용될 수 있음을 보여준다.

본 논문의 구성은 다음과 같다. 제 II장에서는 협력통신 네트워크 모델에 대해서 분석한다. 제 III장에서는 Q-Learning과 제안된 릴레이 선택 기법에 대해서 서술한다. 제 IV장에서는 제안된 기법의 모의실험 결과에 대해서 분석하고, 마지막으로 제 V장에서는 본 논문의 결론을 맺는다.

II. 협력통신 네트워크 모델

1. Amplify-and-Forward 릴레이에 기반을 둔 협력통신에서 MRC 출력의 수학적 분석

본 논문에서는 그림 1과 같이 N 개 릴레이의 도움을 받아서 목적지로 정보를 전송하는 무선 협력통신 네트워크를 고려한다. 소스와 릴레이 사이의 채널과 릴레이와 목적지 사이의 채널은 각각 독립적이라고 가정한다. 협력통신은 소스의 정보를 전달하기 위해서 두 번의 과정을 거친다. 첫 번째 과정에서, 소스는 목적지와 각 릴레이들에게 자신의 정보를 전달한다. 두 번째 과정에서, 각 릴레이들은 전송받은 소스의 정보를 증폭하여 목적

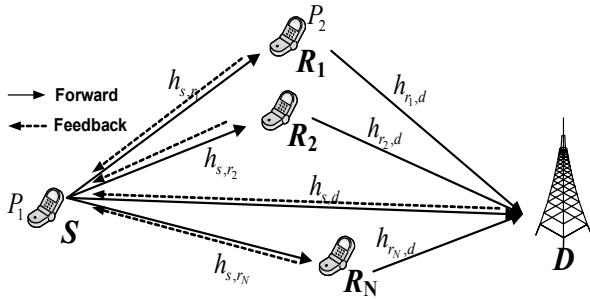


그림 1. 릴레이를 이용하는 협력통신 네트워크 시스템
Fig. 1. A cooperative relaying network system.

지로 그 정보를 전달한다.

첫 번째 과정에서, 릴레이 R_i 와 목적지가 소스로부터 받은 정보는 다음 수식과 같다.

$$y_{s,i} = \sqrt{P_s} h_{s,i} x + \eta_{s,i}, \quad i = 1, \dots, N \quad (1)$$

$$y_{s,d} = \sqrt{P_s} h_{s,d} x + \eta_{s,d} \quad (2)$$

여기서, P_s 는 소스에서의 송신 전력, x 는 전송된 정보 심벌을 의미한다. 그리고 $\eta_{s,d}$ 와 $\eta_{s,i}$ 는 각각 목적지와 릴레이 R_i 에서의 가우시안 랜덤변수로 모델링 되는 잡음 성분을 의미한다. 또한 $h_{s,d}$ 와 $h_{s,i}$ 는 각각 소스-목적지 채널과 소스-릴레이 채널의 채널 계수를 의미한다.

두번째 과정에서, 릴레이 R_i 는 소스로부터 받은 신호를 증폭하고 목적지에 증폭된 신호를 재전송한다. 릴레이 R_i 로부터 목적지에서 받은 신호는 다음 수식과 같다^[13].

$$y_{i,d} = \beta_i h_{i,d} y_{s,i} + \eta_{i,d}, \quad i = 1, \dots, N \quad (3)$$

여기서 β_i 는 전력 제한을 만족시키는 비례 계수로 다음과 같다.

$$\beta_i \leq \frac{\sqrt{P_i}}{\sqrt{P_s |h_{s,i}|^2 + N_0}} \quad (4)$$

위 식에서 P_i 는 릴레이 R_i 의 송신 전력을 의미하고, 채널 계수 $h_{s,d}$, $h_{s,i}$, $h_{i,d}$ 는 분산이 $\delta_{s,d}^2$, $\delta_{s,i}^2$, $\delta_{i,d}^2$ 이고 평균이 0인 복소 가우시안 랜덤변수로 모델링된다. 한편 잡음은 분산이 N_0 이고 평균이 0인 복소 가우시안 랜덤변수로 모델링된다.

수신기에서 순시 채널 이득에 대한 정보를 모두 안다

고 가정하면, MRC 검출기의 출력은 다음 식과 같다.

$$y_d = \alpha_s y_{s,d} + \sum_{i=1}^N \alpha_i y_{i,d} \quad (5)$$

여기서 $\alpha_s = \frac{\sqrt{P_s} h_{s,d}}{N_0}$ 이고 $\alpha_i = \frac{\sqrt{P_s} \beta_i h_{s,i} h_{i,d}}{(\beta_i^2 |h_{i,d}|^2 + 1) N_0}$ 이다.

이러한 다중 릴레이를 이용하는 무선 협력통신 네트워크에서 목적지의 신호대잡음비 (Signal-to-Noise ratio; SNR)는 다음과 같이 표현 된다.

$$SNR_d = \Upsilon_s + \sum_{i=1}^N \Upsilon_i \quad (6)$$

여기서 $\Upsilon_s = P_s |h_{s,d}|^2 / N_0$ 이고 $\Upsilon_i = \frac{P_s \beta_i^2 |h_{i,d}|^2 |h_{s,i}|^2}{(\beta_i^2 |h_{i,d}|^2 + 1) N_0}$ 이다.

알고리즘의 복잡도를 낮추기 위해서, 목적지의 SNR_d 은 다음과 같이 근사화할 수 있다^[4].

$$\begin{aligned} SNR_d &\leq \frac{P_s |h_{s,d}|^2}{N_0} + \frac{P_s}{N_0} \sum_{i=1}^N \frac{P_i |h_{i,d}|^2 |h_{s,i}|^2}{P_i |h_{i,d}|^2 + P_s |h_{s,i}|^2} \\ &= \frac{P_s |h_{s,d}|^2}{N_0} + \frac{1}{N_0} \sum_{i=1}^N \left((P_s |h_{s,i}|^2)^{-1} + (P_i |h_{i,d}|^2)^{-1} \right)^{-1} \\ &\leq \frac{P_s |h_{s,d}|^2}{N_0} + \frac{1}{2N_0} \sum_{i=1}^N \sqrt{P_s P_i} |h_{s,i}| |h_{i,d}| \end{aligned} \quad (7)$$

이러한 결과를 정리하면, L 개 릴레이를 이용하는 협력통신에서 목적지의 총 SNR은 다음과 같다.

$$SNR_d = \frac{P_s |h_{s,d}|^2}{N_0} + \frac{1}{2N_0} \sum_{i=1}^L \sqrt{P_s P_i} |h_{s,i}| |h_{i,d}| \quad (8)$$

2. MRC에 기반을 둔 릴레이 선택 기법

SNR_d 에 기반을 두어 [4]에서는 다음과 같은 릴레이 선택 기법을 제안하였다.

- 협력통신에 모든 릴레이들이 참여한다고 가정하고, 이 릴레이들의 집합을 $\Phi = \{1, \dots, N\}$ 로 정의한다.
- 모든 릴레이들의 인덱스 집합이 다음과 같이 정렬된다고 가정한다.

$$|h_{s,i}| |h_{i,d}| \geq |h_{s,(i+1)}| |h_{(i+1),d}| \quad (9)$$

- 목적지에서의 총 SNR을 증가시키기 위해서, 다

음 식을 만족하도록 한다.

$$SNR_d(N) - SNR_d(N-1) < 0 \quad (10)$$

- 식 (10)에 식 (8)을 대입하여 다음과 같이 간단하게 정리한다.

$$|h_{s,N}| |h_{N,d}| < \frac{2(P'_s - P_s)}{\sqrt{P_s P_i}} |h_{s,d}|^2 \quad (11)$$

$$+ \frac{\sqrt{P'_s P'_i} - \sqrt{P_s P_i}}{\sqrt{P_s P_i}} \sum_{i=1}^{N-1} |h_{s,i}| |h_{i,d}| = A$$

여기서 P'_s 와 P'_i 은 하나의 릴레이가 협력 릴레이 집합에서 제거 되었을 때의 소스와 각 릴레이에서의 송신 전력을 의미한다.

- 모든 릴레이 선택 구간 동안 다음의 알고리즘으로 협력 릴레이 개수를 선택한다.
 - ① 만약 $|h_{s,N}| |h_{N,d}| < A$ 이면, 릴레이 집합 Φ 에서 릴레이 N 을 제외시키고 Φ 를 업데이트하고, 그렇지 않으면 단계 ③으로 이동함
 - ② $N=N-1$ 으로 설정하고, A 를 업데이트함. 만약 $N=1$ 이면 단계 ③으로 이동하고, 그렇지 않으면 단계 ①로 이동함
 - ③ 협력 릴레이 N 개를 선택 후 종료함

제 IV장에서, [4]에서 제안한 이러한 릴레이 선택 기법과 본 논문에서 제안한 Q-Learning을 이용한 릴레이 선택 기법의 통신 성능을 비교한다.

III. Q-Learning을 이용한 제안된 릴레이 선택 기법

1. Q-Learning 이론

Q-Learning은 기본적으로 환경, 에이전트, 상태, 행동, 그리고 보상으로 구성된 강화 학습 알고리즘이다. Q-Learning 알고리즘에서 환경은 유한 상태 이산 시간 확률 시스템 (Finite-State Discrete-Time Stochastic System)으로 구성된다. $S = \{s_1, \dots, s_n\}$ 을 주어진 시스템에서 가능한 모든 상태들의 집합, $A = \{a_1, \dots, a_m\}$ 을 가능한 모든 행동들의 집합이라고 하자. 에이전트는 환경 $s \in S$ 에서 행동 $a \in A$ 를 취한다. 그리고 학습하기 위해서

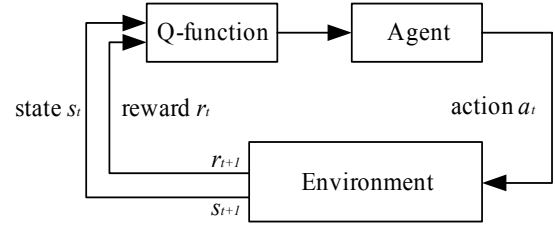


그림 2. Q-Learning에서 에이전트와 환경간의 반복 Fig. 2. Agent-environment iteration in Q-Learning.

그 환경으로부터 두 종류의 피드백을 받는다. 첫째로, 에이전트가 현재 취한 행동으로 인한 즉각적인 보상 $r(s,a)$ 을 받는다. 두번째로, 현재 행동으로 인해서 나타나는 미래 환경에 대한 보상인 Q-Value $Q(s',a')$ 을 받는다. Q-Learning의 핵심은 환경으로부터 이러한 피드백들을 최대한 긍정적으로 받고, 그 후에 Q-Value들로 구성된 Q-Table을 업데이트 하는 것이다. 이 때 기본적인 Q-Function은 다음과 같다.

$$Q(s,a) = r(s,a) + \gamma \max_{a'} \quad (12)$$

여기서, $s' \in S$ 는 현재의 환경 s 에서 행동 a 를 취했을 때 나타나는 다음 환경, a' 은 $s' \in S$ 에서 취할 수 있는 가능한 모든 행동들을 의미한다. 또한, γ ($0 < \gamma < 1$)은 할인 계수 (Discount Factor)로 미래의 상황이 현재의 학습에 얼마나 영향을 미칠 것인가를 결정한다. 즉, Q-Learning은 반복된 학습에 의해서 축적된 보상값을 이용하여 주어진 환경에서 최적의 행동을 찾아 나가게 된다. 이러한 과정은 그림 2와 같이 묘사되어질 수 있다.

2. Q-Learning을 이용한 제안된 릴레이 선택 기법

본 논문에서는 다음과 같이 Q-Learning을 위한 상태, 행동, 그리고 보상을 정의한다.

가. 상태 (State)

본 논문에서는 N 개의 릴레이를 가지고 있는 무선 협력통신 네트워크에서 릴레이 선택을 위한 Q-Learning 알고리즘을 제안하므로, 상태는 $s_i = i$ ($i = 1, \dots, N$)으로 정의한다. 여기서 i 는 협력통신을 위해서 선택된 협력 릴레이들의 수를 의미한다.

나. 행동 (Action)

릴레이 선택을 위해 행동 a 를 행하는 것은 이용 가

능한 N 개 릴레이로부터 L ($\leq N$)개 릴레이를 선택하는 것을 의미한다. 즉, 행동 a 는 $a=n$ ($n \in \{1, \dots, N\}$)로 정의된다. 결과적으로, 행동은 L_1 ($\leq N$)개 릴레이를 사용하는 상태에서 L_2 ($\leq N$)개 릴레이를 사용하는 상태로 이동하는 것을 의미한다. 이러한 행동을 취함으로써 인해서 SER 성능이 좋아지거나 또는 나빠지는 것은 현재 상태에서는 고려하지 않는다. 이것은 적절한 보상이 현재 행동을 위해서 주어짐으로써, 시간이 지남에 따라 에이전트에 의해 학습된다.

다. 보상 (Reward)

Q-Learning을 위해서 보상 r 을 정의하는 것은 중요한 이슈이다. 본 논문에서는 적은 개수의 협력 릴레이들을 이용하면서, 동시에 우수한 SER 성능을 얻기 위하여 r 을 다음 식과 같이 정의한다.

$$r = \frac{SNR_d(state_{next}) - SNR_d(state_{current})}{|state_{next} - state_{current}|} \quad (13)$$

여기서 $SNR_d(state_{next})$ 과 $SNR_d(state_{current})$ 는 각각 현재 상태와 행동을 취한 후의 다음 상태에서 MRC 출력의 SNR을 의미한다. 이 식에서 $SNR_d(state_{current})$ 가 $SNR_d(state_{next})$ 보다 작다면 부정적인 보상 받을 것이고, 그 반대의 경우에는 긍정적인 보상을 받을 것이다. 또한, $|state_{next} - state_{current}|$ 향으로 다음 상태와 현재 상태에서의 SNR_d 의 차이를 나누는 것에 의해서, 단순히 높은 SNR_d 를 갖는 릴레이 개수를 선택했을 때 높은 보상을 받기 보다는 적은 릴레이 개수로 적절한 SNR_d 를 갖는 상태로 이동하도록 행동을 취했을 때 높은 보상을 받도록 하였다.

본 논문에서는 식 (13)에서 정의된 보상 r 이외에 다음 식과 같이 보상 r 을 추가로 정의하였다.

$$r = \frac{SNR_d(state_{next}) - SNR_d(state_{current})}{|state_{next} - state_{current}|} \times state_{next} \quad (14)$$

여기서는 식 (13)에 $state_{next}$ 를 곱해줌으로써 식 (13)에 비해서 긍정적인 보상과 부정적인 보상에 대한 격차를 더 크게 주었다.

릴레이 선택을 위해서 제안된 Q-Learning 알고리즘은 다음과 같이 동작한다.

- 협력통신에 모든 릴레이들이 참여하고, 모든 릴레이들의 인덱스 집합은 [4]에서 제안한 것과 같은 방법으로 정렬된다고 가정한다.
- Q-Table을 초기화하고, 초기 상태를 랜덤하게 선택한다.
- 모든 릴레이 선택 구간 동안 다음 알고리즘으로 협력 릴레이 개수를 선택한다.
 - ① 현재 상태에서 가능한 모든 행동들로부터 하나의 행동을 선택함
 - ② 선택된 행동을 사용해서, 다음 상태 s' 으로 이동함
 - ③ 가능한 모든 행동들에 기반하여, 다음 상태 s' 가운데 Q-Value의 최대값을 가지는 행동 a' 을 선택함
 - ④ 식 (12)를 사용해서 Q-Value를 업데이트 함
 - ⑤ Q-Table에서 최대의 Q-Value를 가지는 협력 릴레이 개수를 선택함
 - ⑥ 다음 상태 s' 을 현재 상태로 설정함

위와 같은 릴레이 선택 과정에서, 제안 기법은 단순히 Q-Learning을 위한 행동을 취하고 그 행동에 대한 보상을 받은 후 Q-Table을 업데이트하게 된다. 또한, 제안 기법에서 보상은 식 (8)을 두 번만 계산하면 되므로 알고리즘적인 복잡도와 계산량이 크게 감소한다. 반면에 다음 장의 모의실험 결과를 통해 알 수 있듯이 수학적 분석을 통한 릴레이 선택 기법과 비교했을 때 제안 기법은 적은 개수의 릴레이를 사용하지만 거의 유사한 SER 성능을 얻는 것을 보인다.

IV. 모의실험 결과

본 장에서는 Q-Learning을 이용하는 제안된 릴레이 선택 알고리즘의 성능을 모의실험을 통해 검증한다. 통신을 위해서 QPSK (Quadri-Phase Shift Keying) 변조 방식을 사용하였고 비트 에너지를 1로 정규화 하였다. 소스와 목적지 사이의 채널 계수는 1로 고정되어 있고, 소스와 각 릴레이 사이의 채널 계수와 각 릴레이와 목적지 사이의 채널 계수는 [1,10]의 균등 분포를 가지고 있다고 가정하였다. 잡음 분산은 1로 정규화 하였으며, 하나의 전송 블록에서 채널 계수는 변하지 않으나 서로 다른 블록 사이의 채널 계수는 독립적으로 변한다고 가

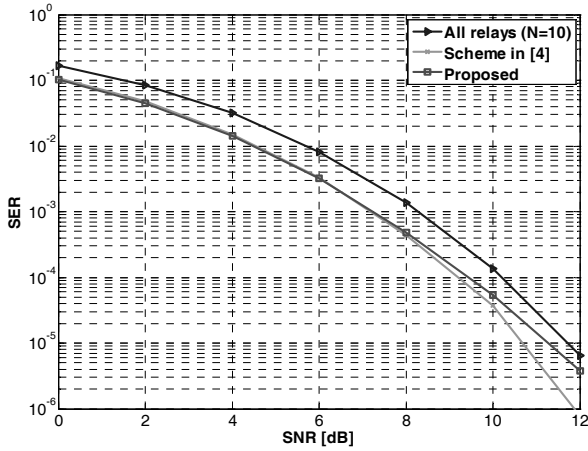


그림 3. 릴레이 개수 $N=10$ 일 때 다양한 기법에 대한 SER 성능

Fig. 3. SER performance for various schemes with $N=10$ relays.

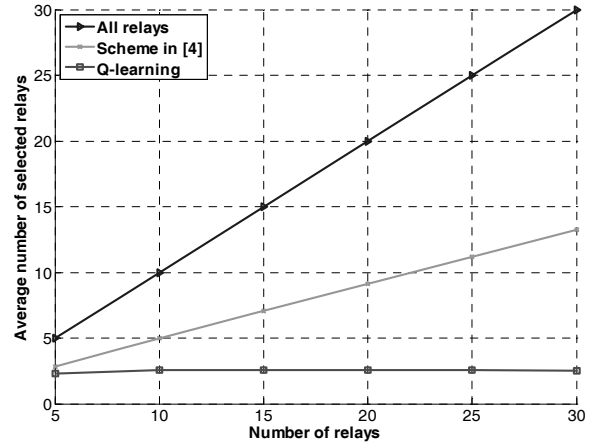


그림 5. SNR = 10 dB일 때 다양한 기법에서 선택된 릴레이의 평균 개수 비교

Fig. 5. Comparison of average number of selected relays at SNR = 10 dB.

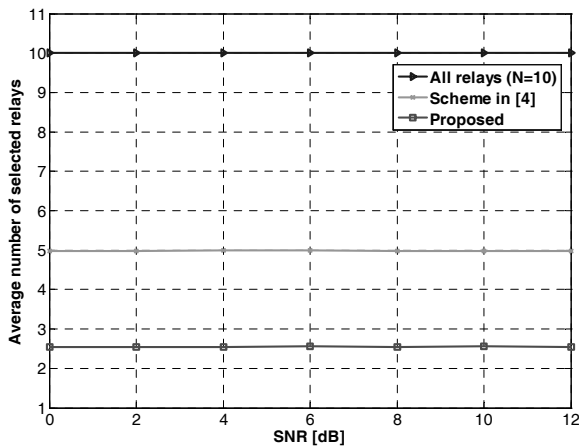


그림 4. $N=10$ 개의 릴레이 중에서 선택된 평균 릴레이 개수

Fig. 4. The average number of selected relays among $N=10$ relays.

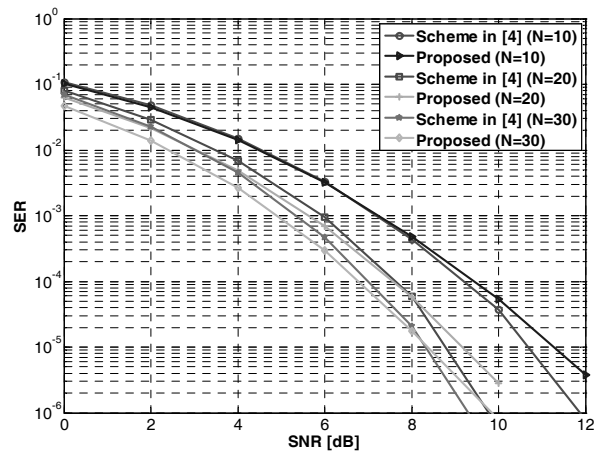


그림 6. 협력통신에 참여할 수 있는 다양한 릴레이 개수에 대한 SER 성능

Fig. 6. SER performance for various relay numbers that can participate in cooperative communication.

정하였다. 협력통신 네트워크에서 중요하게 다루어야 하는 송신 전력의 경우, 비교되는 릴레이 선택 기법과 제안된 릴레이 선택 기법의 복잡도를 낮추기 위해서 본 논문에서는 동등 전력 할당을 사용하였다. 동등 전력 할당은 전체 전력 P 의 절반은 소스와 선택되었던 협력 릴레이 집합에 각각 할당되고, 협력 릴레이 집합에 할당된 전체 전력 P 의 절반은 다시 선택된 각 협력 릴레이들에게 동등하게 분배되는 것을 가정한다. 그림 3부터 그림 5까지 보이는 실험 결과는 이용 가능한 전체 릴레이 개수가 $N=10$ 임을 가정한 것이고, 그림 3부터 그림 6까지 보이는 실험 결과는 식 (13)의 보상 방안을

이용한 것이다.

그림 3은 모든 릴레이가 참여할 때, [4]에서 제안한 릴레이 선택 기법이 사용되었을 때, 그리고 Q-Learning을 이용한 제안된 릴레이 선택 기법이 사용되었을 때의 SER 성능을 보여준다. 그림 3에서 SNR = 8 dB까지 본 논문에 제안된 기법과 [4]에서 제안한 기법의 SER 성능이 매우 유사하고, 그 이후에는 제안 기법의 SER 성능이 약간 악화되는 것을 관찰 할 수 있다.

그림 4는 협력통신을 위해서 사용된 평균 릴레이 개수를 보여준다. 여기서, 본 논문에서 제안된 기법이 [4]에서 제안한 기법 보다 약 2.5개의 릴레이를 절약하고,

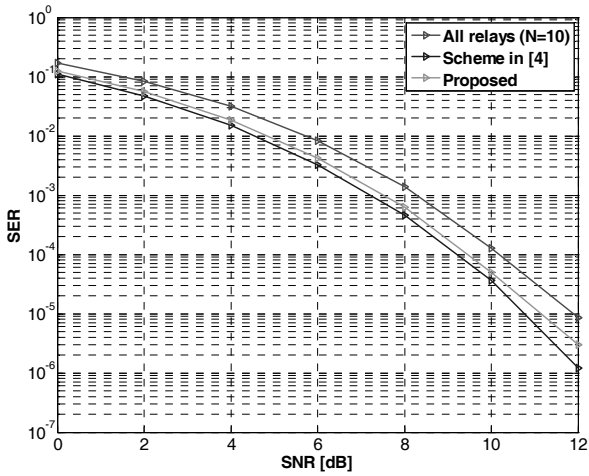


그림 7. 식 (14)의 보상 방안 이용시 릴레이 개수 $N=10$ 일 때 다양한 기법에 대한 SER 성능
Fig. 7. SER performance for various schemes with $N=10$ relays when using a reward in (14).

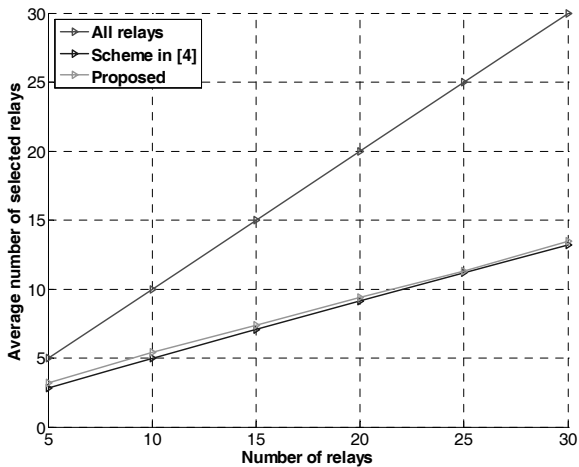


그림 8. 식 (14)의 보상 방안 이용시 SNR = 10 dB일 때 다양한 기법에서 선택된 릴레이의 평균 개수 비교
Fig. 8. Comparison of average number of selected relays at SNR = 10 dB when using a reward in (14).

모든 릴레이가 협력통신에 참여 했을 때 보다 약 7.5개의 릴레이를 더 적게 사용하는 것을 확인 할 수 있다.

그림 5는 SNR = 10 dB에서 이용 가능한 릴레이 개수가 증가함에 따라서 협력통신을 위해 선택된 평균 릴레이 개수를 보여준다. 이 결과로부터, [4]에서 제안한 릴레이 선택 기법과 다르게 본 논문에서 제안된 기법은 이용 가능한 릴레이 개수가 증가함에도 불구하고 평균적으로 약 2.5개의 고정된 개수의 릴레이를 선택하는 것을 보여준다. 그림 6은 협력통신에 참여하는 다양한 릴레이 개수

에 따른 SER 성능을 보여준다. 그림 5와 6에서 볼 수 있듯이, 참여 가능한 릴레이 개수가 증가함에 따라 선택되는 협력 릴레이 개수 또한 증가하는 것이 SER 성능을 증가시키기 위해서 합리적인 접근이다. 하지만, 실험 결과를 통해서 SNR이 8 dB 보다 작을 때, 협력 릴레이 개수가 약간 감소하는 것은 SER 성능에 거의 영향을 미치지 못하고, 더 나아가서 SNR이 8 dB 이상일 때는 비록 SER 성능이 약간 저하되지만 협력 릴레이 개수를 급격하게 줄일 수 있다는 것을 보여준다.

그림 7과 8은 식 (14)의 보상 방안을 이용했을 때에 대한 실험 결과를 보여준다. 앞서 언급한 바와 같이, 식 (14)는 식 (13)에 $state_{next}$ 항을 추가함으로써 긍정적인 보상과 부정적인 보상의 격차를 크게 만든 것이다. 그림 7과 8에서 볼 수 있듯이, 식 (14)의 보상을 이용한 릴레이 선택 기법은 [4]에서 제안한 기법과 비교했을 때 SER 성능과 협력 릴레이 개수에 대한 측면에서 매우 비슷한 성능을 보인다. 하지만 식 (13)을 사용할 때와 비교해서 보상의 정의만 바뀌었으므로, 여전히 수학적 분석을 통해서 릴레이를 선택하는 [4]의 기법에 비해서 알고리즘적인 복잡도와 계산량은 적다. 이러한 실험 결과는 Q-Learning에서 보상 r 을 정의하는 형태에 따라서 시스템의 목적을 다르게 설정할 수 있음을 보여준다.

지금까지의 실험 결과들은 제안된 기법이 행동에 대한 보상을 주고 Q-Table을 업데이트하는 간단한 릴레이 선택 과정과 함께 협력 릴레이 개수를 급격하게 감소시키지만 우수한 SER 성능을 얻을 수 있음을 보여준다.

V. 결 론

본 논문은 무선 협력통신 네트워크에서 Q-Learning을 이용하는 릴레이 선택 기법을 제안하였다. 제안된 기법에서, Q-Learning을 위하여 상태, 행동, 보상을 적절히 정의 하는 것에 의해서 복잡한 수학적 계산 없이 우수한 SER 성능을 얻는 시스템이 고안되었다. 특히, [4]에서 제안한 기법에서는 이용 가능한 릴레이 개수가 증가함에 따라서 협력 릴레이 개수 또한 선형적으로 증가하였지만, 본 논문에서 제안된 기법은 평균적으로 약 2.5개의 협력 릴레이만을 사용하였다. 또한, [4]는 릴레이 선택을 위해서 최악의 경우 식 (11)을 $N-1$ 번 계산하

지만, 본 논문에서 제안된 기법은 식 (8)을 오직 두 번만 계산한다. 그러므로 본 논문에서 제안된 기법은 복잡도 면에서 매우 우수하다고 할 수 있다. 이러한 결과들은 적은 개수의 협력 릴레이를 사용하는 것에 의해 무선자원 이용의 효율성을 개선시킬 수 있음을 보여준다. 또한, Q-Learning에서 보상 r 을 정의하는 방법에 따라서 시스템의 목적을 다르게 설정 할 수 있음을 보여주었다. 본 논문의 연구 결과는 자가 학습을 하는 Q-Learning이 다수의 릴레이를 사용하는 미래의 무선통신 네트워크를 위해서 효율적으로 사용될 수 있음을 보여준다.

참 고 문 헌

- [1] A. Nosratinia, T. E. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Commun. Mag.*, vol. 42, pp. 74-80, October 2004.
- [2] Y. Zhao, R. S. Adve, and T. J. Lim, "Improving amplify-and-forward relay networks: Optimal power allocation versus selection," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 3114-3123, August 2007.
- [3] Z. Kenan and T. M. Lok, "A relay selection scheme under optimal power allocation," *Proc. ICCS 2008*, pp. 1609-1613, Guangzhou, China, November 2008.
- [4] H. Wu, Y. Wang, C. Xiong, and D. Yang, "A novel relay selection scheme with simplified power allocation for wireless relay networks," *Proc. IEEE GLOBECOM 2009*, pp. 1-5, Honolulu, USA, December 2009.
- [5] F. Ke, S. Feng, and H. Zhuang, "Relay selection and power allocation for cooperative network based on energy pricing," *IEEE Commun. Lett.*, vol. 14, no. 5, pp. 396-398, May 2010.
- [6] L. Song, "Relay selection for two-way relaying with amplify-and-forward protocols," *IEEE Trans. Veh. Technol.*, vol. 60, no. 4, pp. 1954-1959, April 2011.
- [7] R. S. Sutton and A. G. Barto, *Reinforcement Learning*, MIT Press, 1998.
- [8] 김재명, "Cognitive Radio 기술개요 및 발전방향," *대한전자공학회지*, 제36권, 제6호, 20-27쪽, 2009년 6월.
- [9] 조정일, 서종필, 유재호, 정재학, "Cognitive Radio의 네트워크 및 물리계층 연구동향," *대한전자공학회지*, 제36권, 제6호, 39-51쪽, 2009년 6월.
- [10] A. Galindo-Serrano and L. Giupponi, "Distributed Q-learning for aggregated interference control in cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 4, pp. 1823-1834, April 2010.
- [11] O. V. Den Biggelaar, "Sensing time and power allocation for cognitive radios using distributed Q-learning," *EURASIP Jour. Wireless Commun. & Networking*, vol. 2012, no. 138, April 2012.
- [12] J. Peng, J. Li, S. Li, and Jing Li, "Multi-relay cooperative mechanism with Q-learning in cognitive radio multimedia sensor networks," *Proc. IEEE TrustCom-11/IEEE ICSS-11/FCST-11*, pp. 1624-1629, Changsha, China, November 2011.
- [13] K. J. R. Liu, A. K. Sadek, W. Su, and A. Kwasiński, *Cooperative Communication and Networking*, Cambridge University Press, 2009.

저 자 소 개



정 홍 규(학생회원)
2012년 숭실대학교 정보통신전자공학부 학사 졸업.
2012년 3월~현재 숭실대학교 정보통신공학과 석사과정.
<주관심분야 : 무선통신 시스템, 통신신호처리>



신 요 안(평생회원)-교신저자
1987년 2월 서울대학교 전자공학과 졸업 (공학사).
1989년 2월 서울대학교 대학원 전자공학과 졸업 (공학석사).

1992년 12월 Univ. of Texas at Austin 전기및컴퓨터 공학과 졸업 (공학박사).

1992년 12월~1994년 7월 오스틴 소재 MCC (Microelectronics & Computer Technology Corp.) Member of Technical Staff.

1994년 9월~현재 숭실대학교 정보통신전자공학부 전임강사/조교수/부교수/교수, 학부장.

2009년 9월~2010년 8월 Univ. of British Columbia 전기및컴퓨터공학과 방문교수.

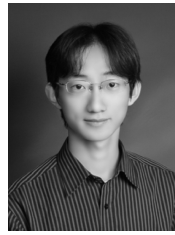
2005년 1월~2006년 12월 한국통신학회 부호및 정보이론연구회 위원장.

2007년 1월~2008년 12월 대한전자공학회 이사.

2008년 1월~2008년 12월 한국통신학회 이동통신연구회 위원장.

2011년 1월~현재 한국통신학회 집행이사.

<주관심분야 : 무선 및 이동통신 시스템, 통신신호처리>



김 광 열(학생회원)
2007년 공주대학교 정보통신공학과 학사 졸업.
2007년~2009년 한국전자통신연구원 위촉연구원.
2009년 9월~현재 숭실대학교 정보통신공학과 석박사통합 과정.

<주관심분야 : 무선통신 시스템, 최적화알고리즘>