



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2014년11월19일
 (11) 등록번호 10-1462199
 (24) 등록일자 2014년11월10일

- | | |
|---|-------------------------|
| (51) 국제특허분류(Int. Cl.)
H04B 7/02 (2006.01) | (73) 특허권자
승실대학교산학협력단 |
| (21) 출원번호 10-2013-0053056 | (72) 발명자
신요안 |
| (22) 출원일자 2013년05월10일
심사청구일자 2013년05월10일 | (74) 대리인
특허법인태백 |
| (65) 공개번호 10-2014-0103797 | 정홍규 |
| (43) 공개일자 2014년08월27일 | 김광열 |
| (30) 우선권주장
1020130017581 2013년02월19일 대한민국(KR) | |
| (56) 선행기술조사문헌
Ana Galindo-Serrano et al.; "Distributed Q-Learning for Aggregated Interference Control in Cognitive Radio Networks", IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, VOL. 59, NO. 4, MAY 2010*
J. Peng et al.; "Multi-relay cooperative mechanism with Q-learning in cognitive radio multimedia sensor network", 2011 International Joint Conference of IEEE TrustCom-11/IEEE ICSS-11/FCST-11*
*는 심사관에 의하여 인용된 문헌
기술이전 희망 : 기술양도, 실시권허여, 기술지도 | |

전체 청구항 수 : 총 4 항

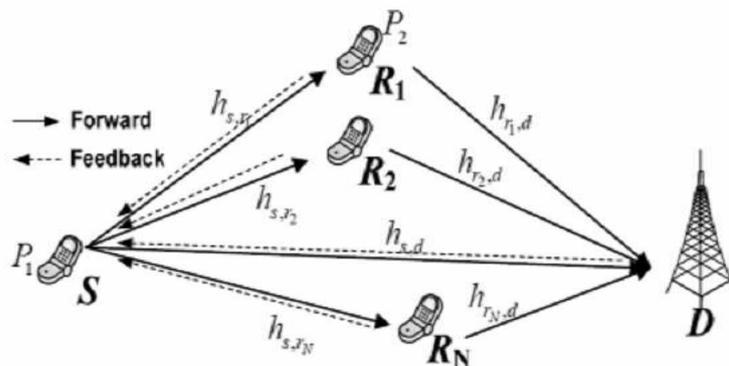
심사관 : 이정수

(54) 발명의 명칭 자율적인 행동 학습(Q-Learning)을 이용한 협력 통신을 위한 릴레이 선택 장치 및 그 방법

(57) 요약

본 발명은 자율적인 행동 학습(Q-Learning)을 이용한 협력 통신을 위한 릴레이 선택 장치 및 그 방법에 관한 것이다. 본 발명에 따른 협력 통신을 위한 릴레이 선택 방법은, 모든 릴레이가 협력 통신에 참여하고, 모든 릴레이의 인덱스 집합을 정의하는 단계; Q-테이블을 초기화하고, 초기 상태를 랜덤하게 선택하는 단계; 및 모든 릴레이(뒷면에 계속)

대표도 - 도1



이 선택 구간 동안 협력 릴레이 개수를 선택하는 단계를 포함하고, 상기 협력 릴레이 개수를 선택하는 단계는 현재 상태에서 가능한 모든 행동들로부터 하나의 행동을 선택하는 단계; 상기 선택된 행동을 이용하여 다음 상태로 이동하는 단계; 가능한 모든 행동들에 기반하여 상기 다음 상태 가운데 Q 값의 최대값을 가지는 행동을 선택하는 단계; 상기 Q 값을 업데이트하고, 상기 Q-테이블에서 최대의 Q 값을 가지는 협력 릴레이 개수를 선택하는 단계; 및 다음 상태를 현재 상태로 선택하는 단계를 포함한다.

이와 같이 본 발명에 따르면, 물리 계층 관점에서 자율적인 행동 학습(Q-Learning)을 위한 상태, 행동, 보상값을 명확하게 정의하고 전체 알고리즘을 단계화할 수 있고, 자율적인 행동 학습(Q-Learning)을 이용한 제안 기법은 수학적 분석을 통해서 릴레이를 선택하는 기법과 비교했을 때 시스템의 복잡도를 효율적으로 저감시키며 유사한 심벌오율(Symbol Error Rate) 성능을 낼 수 있다. 또한, 자율적인 행동 학습(Q-Learning) 알고리즘은 차세대 무선통신 시스템의 성능 향상을 위해서 효과적으로 사용될 수 있다.

이 발명을 지원한 국가연구개발사업

과제고유번호	20120005330
부처명	교육과학기술부
연구관리전문기관	한국연구재단
연구사업명	중견(핵심)연구자지원사업 2차년도
연구과제명	생태계 모방 인지 라디오 스펙트럼 공유 기술
기여율	1/1
주관기관	승실대학교 산학협력단
연구기간	2011.05.01 ~ 2014.04.30

특허청구의 범위

청구항 1

자율적인 행동 학습(Q-Learning)을 이용한 협력 통신을 위한 릴레이 선택 장치에 있어서,

상기 자율적인 행동 학습의 상태, 행동 및 보상에 대해, 선택된 협력 릴레이 개수를 상태로, 상기 릴레이 개수의 변화를 행동으로, 그리고 신호대잡음비(Signal-to-Noise Ratio: SNR)에 관한 함수값을 보상으로 설정하는 설정부;

상기 자율적인 행동 학습의 상태에서 임의의 상태값 및 상기 선택된 상태값에서 가능한 하나 이상의 행동값을 선택하는 선택부;

상기 선택된 상태값 및 행동값을 기초로 이에 대응하는 하나 이상의 보상값을 연산하는 연산부; 및

상기 하나 이상의 보상값 중에서 최고의 보상값을 가지는 행동의 릴레이 개수를 최적의 릴레이 개수로 결정하는 제어부를 포함하되,

상기 연산부는,

아래 수학적식에 따라 보상값을 업데이트 하는 협력 통신을 위한 릴레이 선택 장치:

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

여기서, s는 현재 상태, s'는 미래 상태, a는 현재 행동 그리고 a'는 미래 행동을 그리고 r은 즉각적인 보상값을 의미하고, γ ($0 < \gamma < 1$)는 미래 상황이 현재의 학습에 미치는 영향을 결정하는 할인계수(Discount Factor)이다.

청구항 2

청구항 2은(는) 설정등록료 납부시 포기되었습니다.

제 1항에서,

상기 연산부는,

다음의 수학적식과 같이 상기 보상값을 연산하는 협력 통신을 위한 릴레이 선택 장치:

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

여기서 r은 상기 보상값, SNR_d 는 수신단의 신호대잡음비 그리고 state(상태)는 선택된 릴레이 개수를 각각 의미한다.

청구항 3

청구항 3은(는) 설정등록료 납부시 포기되었습니다.

제 1항에 있어서,

상기 연산부는,

다음의 수학적식과 같이 상기 보상값을 연산하는 협력 통신을 위한 릴레이 선택 장치:

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

여기서 r은 상기 보상값, SNR_d는 수신단의 신호대잡음비 그리고 state(상태)는 선택된 릴레이 개수를 각각 의미한다.

청구항 4

삭제

청구항 5

제 1항에 있어서,

상기 제어부는,

상기 업데이트된 보상값인, Q-Value Q(s, a)으로 Q-Table을 생성하고,

상기 최적의 릴레이 개수 결정에서 상기 Q-Table을 이용하는 협력 통신을 위한 릴레이 선택 장치.

청구항 6

자율적인 행동 학습(Q-Learning)을 이용한 협력 통신을 위한 릴레이 선택 장치의 릴레이 선택 방법에 있어서,

협력통신 네트워크의 이득을 최대화시키기 위해, 자율적인 행동 학습(Q-Learning)을 이용한 릴레이 선택 방법에 있어서,

상기 자율적인 행동 학습의 상태, 행동 및 보상에 대해, 선택된 협력 릴레이 개수를 상태로, 상기 릴레이 개수의 변화를 행동으로, 그리고 신호대잡음비(Signal-to-Noise Ratio: SNR)에 관한 함수값을 보상으로 각각 설정하는 단계;

상기 자율적인 행동 학습의 상태에서 임의의 상태값 및 상기 선택된 상태값에서 가능한 하나 이상의 행동값을 선택하는 단계;

상기 선택된 상태값 및 행동값을 기초로 이에 대응하는 하나 이상의 보상값을 연산하는 단계; 및

상기 하나 이상의 보상값 중에서 최고의 보상값을 가지는 행동의 릴레이 개수를 최적의 릴레이 개수로 결정하는 단계를 포함하되,

상기 연산하는 단계에서,

다음의 수학적식에 따라 보상값을 업데이트 하는 협력 통신을 위한 릴레이 선택 방법:

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

여기서, s는 현재 상태, s'는 미래 상태, a는 현재 행동, a'는 미래 행동을 그리고 r은 즉각적인 보상값을 의미하고, γ ($0 < \gamma < 1$)는 미래 상황이 현재의 학습에 미치는 영향을 결정하는 할인계수(Discount Factor)이다.

청구항 7

청구항 7은(는) 설정등록료 납부시 포기되었습니다.

제 6항에서,

상기 보상값은,

다음의 수학적식과 같이 연산되는 협력 통신을 위한 릴레이 선택방법:

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

여기서 r은 상기 보상값, SNR_d는 수신단의 신호대잡음비 그리고 state(상태)는 선택된 릴레이 개수를 각각 의미한다.

청구항 8

청구항 8은(는) 설정등록료 납부시 포기되었습니다.

제 6항에 있어서,

상기 보상값은,

다음의 수학적식과 같이 연산되는 협력 통신을 위한 릴레이 선택방법:

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

여기서 r은 상기 보상값, SNR_d는 수신단의 신호대잡음비 그리고 state(상태)는 선택된 릴레이 개수를 각각 의미한다.

청구항 9

삭제

청구항 10

제 6항에 있어서,

상기 최적의 릴레이 개수로 결정하는 단계에서,

상기 업데이트된 보상값인, Q-Value Q(s, a)으로 Q-Table을 생성하고,

상기 최적의 릴레이 개수 결정에서 상기 Q-Table을 이용하는 협력 통신을 위한 릴레이 선택 방법.

명세서

기술분야

[0001] 본 발명은 자율적인 행동 학습(Q-Learning)을 이용한 협력 통신을 위한 릴레이 선택 방법에 관한 것으로서, 보다 상세하게는 협력통신 시스템에서 무선 자원을 효율적으로 사용할 수 있는 자율적인 행동 학습(Q-Learning)을 이용한 릴레이 선택 장치 및 그 방법에 관한 것이다.

배경기술

[0002] 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. 그 중에서도, 릴레이가 소스로부터 전달받은 신호를 단순히 증폭시켜서 목적지로 재전송하는 AF (Amplify-and-Forward) 방식과 소스로부터 전달받은 신호를 복호 후에 전송하는 DF (Decode-and-Forward) 방식이 협력통신을 위해서 주로 사용된다.

[0003] 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.

[0004] 하지만 이러한 수학적 분석은 결과적으로 이상적인 최적의 심별오율 (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.

[0005] 이때, Q-Learning은 기계 학습 분야에서 강화 학습(Reinforcement Learning)에 속하는 알고리즘 중의 하나로, 최근에 또 하나의 차세대 무선통신 시스템으로 주목받고 있는 인지 무선통신(Cognitive Radio)에 적용하는 연구가 많이 진행되고 있다. 특히, 다중 사용자와 다중 채널 인지 무선통신 시스템에서 이차 사용자 사이의 협상 없이 채널을 선택하기 위해서 Q-Learning이 사용되었고, 인지 무선통신에서 사용자들의 스펙트럼 센싱 시간을 공유하고 전력을 효율적으로 할당하기 위해서 분산된 Q-Learning 알고리즘이 사용되었다. 즉, 인지 무선통신 기술을 사용하는 멀티미디어 센서 네트워크에서 협력 릴레이를 선택하기 위하여 Q-Learning이 사용되었다.

[0006] 하지만, 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.

[0007] 본 발명의 배경이 되는 기술은 대한민국 공개특허공보 제10-2007-0071917호(2007.07.04 공개)에 개시되어 있다.

발명의 내용

해결하려는 과제

[0008] 따라서 본 발명이 이루고자 하는 기술적 과제는 협력통신 시스템에서 무선자원을 효율적으로 사용할 수 있는 자율적 행동 학습(Q-Learning)을 이용한 릴레이 선택 장치 및 그 방법을 제공하는 것이다.

과제의 해결 수단

[0009] 이러한 기술적 과제를 이루기 위한 본 발명의 하나의 실시예에 따른 자율적인 행동 학습(Q-Learning)을 이용한 협력 통신을 위한 릴레이 선택 장치는, 상기 자율적인 행동 학습의 상태, 행동 및 보상에 대해, 선택된 협력 릴레이 개수를 상태로, 상기 릴레이 개수의 변화를 행동으로, 그리고 신호대잡음비(Signal-to-Noise Ratio: SNR)에 관한 함수값을 보상으로 설정하는 설정부; 상기 자율적인 행동 학습의 상태에서 임의의 상태값 및 상기 선택된 상태값에서 가능한 하나 이상의 행동값을 선택하는 선택부; 상기 선택된 상태값 및 행동값을 기초로 이에 대응하는 하나 이상의 보상값을 연산하는 연산부; 및 상기 하나 이상의 보상값 중에서 최고의 보상값을 가지는 행동의 릴레이 개수를 최적의 릴레이 개수로 결정하는 제어부를 포함한다.

[0010] 또한, 상기 연산부는, 다음의 수학적식과 같이 상기 보상값을 연산할 수 있다.

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

[0011] 여기서 r은 상기 보상값, SNR_d는 수신단의 신호대잡음비 그리고 state(상태)는 선택된 릴레이 개수를 각각 의미한다.

[0012]

[0013] 또한, 상기 연산부는, 다음의 수학적식과 같이 상기 보상값을 연산할 수 있다.

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

[0014] 여기서 r은 상기 보상값, SNR_d는 수신단의 신호대잡음비 그리고 state(상태)는 선택된 릴레이 개수를 각각 의미한다.

[0015]

[0016] 또한, 상기 연산부는, 아래 수학적식에 의하여 보상값을 업데이트 할 수 있다.

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

[0017]

[0018] 여기서, s는 현재 상태, s'는 미래 상태, a는 현재 행동, a'는 미래 행동을 그리고 r은 즉각적인 보상값을 의미하고, γ ($0 < \gamma < 1$)는 미래 상황이 현재의 학습에 미치는 영향을 결정하는 할인계수(Discount Factor)이다.

[0019] 또한, 상기 제어부는, 상기 업데이트된 보상값인, Q-Value $Q(s, a)$ 으로 Q-Table을 생성하고, 상기 최적의 릴레이 개수 결정에서 상기 Q-Table을 이용할 수 있다.

[0020] 또한, 본 발명의 하나의 실시예에 따른 자율적인 행동 학습(Q-Learning)을 이용한 협력 통신을 위한 릴레이 선택 방법은, 상기 자율적인 행동 학습의 상태, 행동 및 보상에 대해, 선택된 협력 릴레이 개수를 상태로, 상기 릴레이 개수의 변화를 행동으로, 그리고 신호대잡음비(Signal-to-Noise Ratio: SNR)에 관한 함수값을 보상으로 각각 설정하는 단계; 자율적인 행동 학습의 상태에서 임의의 상태값 및 상기 선택된 상태값에서 가능한 하나 이상의 행동값을 선택하는 단계; 상기 선택된 상태값 및 행동값을 기초로 이에 대응하는 하나 이상의 보상값을 연산하는 단계; 및 상기 하나 이상의 보상값 중에서 최고의 보상값을 가지는 행동의 릴레이 개수를 최적의 릴레이 개수로 결정하는 단계; 를 포함한다.

발명의 효과

[0021] 이와 같이 본 발명에 따르면, 물리 계층 관점에서 Q-Learning을 위한 상태, 행동, 보상값을 명확하게 정의하고 전체 알고리즘을 단계화할 수 있고, 수학적 분석을 통해서 릴레이를 선택하는 기법과 비교했을 때 시스템의 복잡도를 효율적으로 저감시키며 유사한 심벌오율(Symbol Error Rate) 성능을 낼 수 있다. 또한, Q-Learning 알고리즘은 차세대 무선통신 시스템의 성능 향상을 위해서 효과적으로 사용될 수 있다.

도면의 간단한 설명

[0022] 도 1은 본 발명의 실시예에 따른 릴레이를 이용하는 협력 통신 네트워크 시스템을 설명하기 위한 구성도이다.
 도 2는 Q-Learning에서 에이전트와 환경간의 반복을 설명하기 위한 구성도이다.
 도 3은 본 발명의 실시예에 따른 릴레이 선택 장치의 구성도이다.
 도 4는 본 발명의 실시예에 따른 릴레이 선택 방법의 흐름도이다.
 도 5 내지 도 10은 본 발명의 실시예에 따른 Q-Learning을 이용한 릴레이 선택 장치 및 그 방법에 대한 실험 결과를 나타낸 그래프이다.

발명을 실시하기 위한 구체적인 내용

[0023] 이하에서는 첨부한 도면을 참조하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본 발명의 실시예를 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[0024] 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 소자를 사이에 두고 "전기적으로 연결"되어 있는 경우도 포함한다. 또한 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미하며, 하나 또는 그 이상의 다른 특징이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0025] 이하 첨부된 도면을 참고하여 본 발명을 상세히 설명하기로 한다.

[0026] 도 1은 본 발명의 하나의 실시예에 따른 릴레이를 이용하는 협력 통신 네트워크 시스템을 설명하기 위한 구성도이다.

[0027] 도 1을 참조하면, 본 발명의 실시예는 N개 릴레이(R_1, R_2, \dots, R_N)의 도움을 받아서 목적지(D)로 정보를 전송하는 무선 협력통신 네트워크를 나타낸다. 소스(S)와 릴레이(R_1, R_2, \dots, R_N) 사이의 채널($h_{s,r1}, h_{s,r2}, \dots, h_{s,rN}$)과 릴레이(R_1, R_2, \dots, R_N)와 목적지(D) 사이의 채널($h_{r1,d}, h_{r2,d}, \dots, h_{rN,d}$)은 각각 독립적이라고 가정한다.

[0028] 협력통신은 소스의 정보를 전달하기 위해서 두 번의 과정을 거친다. 첫번째 과정에서, 소스(S)는 목적지(D)와 각 릴레이(R_1, R_2, \dots, R_N)들에게 자신의 정보를 전달한다. 두번째 과정에서, 각 릴레이(R_1, R_2, \dots, R_N)들은 전송받은 소스의 정보를 증폭하여 목적지(D)로 그 정보를 전달한다.

[0029] 첫번째 과정에서, 릴레이 R_i 와 목적지가 소스로부터 받은 정보는 각각 다음 수학적 식 1 및 수학적 식 2와 같다.

수학적 식 1

[0030]
$$y_{s,i} = \sqrt{P_s} h_{s,i} x + \eta_{s,i}, \quad i = 1, \dots, N$$

수학적 식 2

[0031]
$$y_{s,d} = \sqrt{P_s} h_{s,d} x + \eta_{s,d}$$

[0032] 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 모델링되는 잡음 성분을 의미한다. 또한 $h_{s,d}$ 와 $h_{s,i}$ 는 각각 소스-목적지 채널과 소스-릴레이 채널의 채널 계수를 의미한다.

[0033] 두번째 과정에서, 릴레이 R_i 는 소스로부터 받은 신호를 증폭하고 목적지에 증폭된 신호를 재전송한다. 릴레이 R_i 로부터 목적지에서 받은 신호는 다음의 수학적 식 3과 같다.

수학적 식 3

[0034]
$$y_{i,d} = \beta_i h_{i,d} y_{s,i} + \eta_{i,d}, \quad i = 1, \dots, N$$

[0035] 여기서 β_i 는 전력 제한을 만족시키는 비례 계수로서 다음의 수학적 식 4와 같다.

수학적 식 4

[0036]
$$\beta_i \leq \frac{\sqrt{P_i}}{\sqrt{P_s |h_{s,i}|^2 + N_0}}$$

[0037] 수학적 식 4에서 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.

[0038] 수신기에서 순시 채널 이득에 대한 정보를 모두 안다고 가정하면, MRC 검출기의 출력은 다음의 수학적 식 5와 같다.

수학식 5

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

[0039]

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

[0040]

여기서, 이고, 이다.

[0041]

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

수학식 6

$$SNR_d = \Upsilon_s + \sum_{i=1}^N \Upsilon_i$$

[0042]

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

[0043]

여기서, 이고, 이다.

[0044]

알고리즘의 복잡도를 낮추기 위해서, 목적지의 SNR_d은 다음과 같이 근사화할 수 있다.

수학식 7

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

[0045]

[0046]

이러한 결과를 정리하면, L개 릴레이를 이용하는 협력통신에서 목적지의 총 신호대잡음비(SNR_d)는 다음과 같다.

수학식 8

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

[0047]

[0048] SNR_d을 이용하며 MRC(Maximal-Ratio Combining)에 기반을 두는 릴레이 선택 방법은 아래와 같다.

[0049] 협력통신에 모든 릴레이들이 참여한다고 가정하고, 이 릴레이들의 집합을 $\phi = \{1, \dots, N\}$ 로 정의한다.

[0050] 모든 릴레이들의 인덱스 집합이 아래 수학식 9와 같이 정렬된다고 가정한다.

수학식 9

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

[0051]

[0052] 목적지에서의 총 SNR을 증가시키기 위해서, 다음 수학식 10을 만족하도록 한다.

수학식 10

$$SNR_d(N) - SNR_d(N-1) < 0$$

[0053]

[0054] 상기 수학식 10에 수학식 8을 대입하여 아래 수학식 11과 같이 간단하게 정리한다.

수학식 11

$$|h_{s,N}| |h_{N,d}| < \frac{2(P'_s - P_s)}{\sqrt{P_s P_i}} |h_{s,d}|^2 + \frac{\sqrt{P'_s P'_i} - \sqrt{P_s P_i}}{\sqrt{P_s P_i}} \sum_{i=1}^{N-1} |h_{s,i}| |h_{i,d}| = A$$

[0055]

[0056] 여기서 P'_s와 P'_i은 하나의 릴레이가 협력 릴레이 집합에서 제거되었을 때의 소스와 각 릴레이에서의 송신 전력을 의미한다.

[0057] 모든 릴레이 선택 구간 동안 다음의 알고리즘으로 협력 릴레이 개수를 선택한다.

[0058] ① 만약 $|h_{s,N}| |h_{N,d}| < A$ 이면, 릴레이 집합 ϕ 에서 릴레이 N을 제외시키고 ϕ 를 업데이트하고, 그렇지 않으면 단계 ③으로 이동한다.

[0059] ② $N=N-1$ 로 설정하고, A를 업데이트한다. 만약 $N=1$ 이면 단계 ③으로 이동하고, 그렇지 않으면 단계 ①로 이동한다.

[0060] ③ 협력 릴레이 N개를 선택 후 종료한다.

[0061] 이하에서는, 자율적인 행동 학습(Q-Learning)에 대해 설명한다.

[0062] 도 2는 자율적인 행동 학습(Q-Learning)에서 에이전트와 환경간의 반복을 설명하기 위한 구성도이다. 도 2에 나타난 에이전트(Agent)는 본 발명의 실시예에서는 도 3에서 후술할 릴레이 선택 장치(300)로 구현될 수 있다.

[0063] 도 2를 참조하면, 자율적인 행동 학습(Q-Learning)은 기본적으로 환경(Environment), 에이전트(Agent), 상태

(state, s_t), 행동(action, a_t), 그리고 보상값(reward, r_t)으로 구성된 강화 학습 알고리즘이다.

[0064] 자율적인 행동 학습 (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, $A=\{a_1, \dots, a_n\}$ 을 가능한 모든 행동들의 집합이라고 가정하면, 에이전트는 환경 $s \in S$ 에서 행동 $a \in A$ 를 취한다. 그리고 학습하기 위해서 그 환경으로부터 두 종류의 피드백을 받는다.

[0065] 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 다음의 수학적 식 12와 같다.

수학적 식 12

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

[0066]

[0067] 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 도 2와 같이 도시될 수 있다.

[0068] 이하 본 발명의 자율적인 행동 학습(Q-Learning)을 이용한 협력 통신을 위한 릴레이 선택 장치에 대하여 설명한다.

[0069] 도 3은 본 발명의 실시예에 따른 협력 통신을 위한 릴레이 선택 장치(300)의 구성도이다.

[0070] 도 3을 참조하면, 자율적인 행동 학습(Q-Learning)을 이용한 협력 통신을 위한 릴레이 선택 장치(300)는, 설정부(310), 선택부(320), 연산부(330) 및 제어부(340)를 포함한다.

[0071] 먼저 설정부(310)는 협력 통신을 위한 릴레이 선택 방법에 대응하도록, 자율적인 행동 학습(Q-Learning)의 상태(s), 행동(a) 및 보상(r)을 각각 설정한다.

[0072] 이하 설정부(310)의 "Q-Learning 상태(s)" 설정에 대해 설명한다.

[0073] 본 발명의 실시예에서는 N 개의 릴레이(R_1, R_2, \dots, R_N)를 가지고 있는 무선 협력통신 네트워크에서 릴레이 선택을 위한 Q-Learning 알고리즘을 제안하므로, 상태는 $s_i = i$ ($i=1, \dots, N$)으로 정의한다. 따라서 설정부(310)는 "Q-Learning 상태"를 협력통신을 위해서 선택된 협력 릴레이들의 수로 설정한다.

[0074] 이하 설정부(310)의 "Q-Learning 행동(a)"에 대해 설명한다.

[0075] 자율적인 행동 학습(Q-Learning)을 본 발명에 적용해 보면, 릴레이 선택을 위해 행동 a 를 행하는 것은 이용 가능한 N 개 릴레이로부터 L ($L \leq N$)개 릴레이를 선택하는 것을 의미한다. 즉, 행동 a 는 $a = n$ ($n \in \{1, \dots, N\}$)로 정의된다. 결과적으로, 설정부(310)는 "Q-Learning 행동"을 L_1 ($L_1 \leq N$)개 릴레이를 사용하는 상태에서 L_2 ($L_2 \leq N$)개 릴레이를 사용하는 상태로 이동하는 것, 즉 "릴레이 개수의 변화"로 설정한다.

[0076] 이하 설정부(310)의 "Q-Learning 보상(r)" 설정에 대해 설명한다.

[0077] 본 발명에서 수학적 식 12로 정의되는 Q-Value 값의 기초가 되는 자율적인 행동 학습(Q-Learning)의 보상 r 을 설정하는 것은 본 발명의 특징 중의 하나이다.

[0078] 본 발명에서 설정부(310)는 적은 개수의 협력 릴레이들을 이용하면서, 동시에 우수한 SER 성능을 얻기 위하여 r 을 다음 수학적 식 13과 같이 설정한다.

수학식 13

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

[0079]

[0080]

여기서 $SNR_d(state_{current})$ 와 $SNR_d(state_{next})$ 는 상기 수학식 8에 의해 정의되며, 현재 상태(current)와 행동을 취한 후의 미래 상태(next)에서의 MRC 기반의 릴레이 선택 방법의 신호대잡음비(Signal-to-Noise Ratio: SNR)를 의미한다. 이 식에서 현재의 신호대잡음비($SNR_d(state_{current})$)가 미래의 신호대잡음비($SNR_d(state_{next})$)보다 작다면 보상값은 양의 값이 되고, 그 반대로 현재의 신호대잡음비가 미래의 신호대잡음비보다 크다면 보상값은 음의 값이 된다.

[0081]

또한, 수학식 13에서 $|state_{next} - state_{current}|$ 항으로 미래 상태와 현재 상태에서의 SNR_d 의 차이($SNR_d(state_{next}) - SNR_d(state_{current})$)를 나누는 것에 의해서, 단순히 높은 SNR_d 를 갖는 릴레이 개수를 선택했을 때 높은 보상을 받기보다는 적은 릴레이 개수로 적절한 SNR_d 를 갖는 상태로 이동하도록 행동을 취했을 때 높은 보상값을 가질 수 있다.

[0082]

본 발명의 실시예에서, 설정부(310)는 상기 수학식 13에서 정의된 보상값 r 이외에 다음의 수학식 14와 같이 보상값 r 을 추가로 설정한다.

수학식 14

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

[0083]

[0084]

여기서는 상기 수학식 13에 미래의 상태값($state_{next}$)을 곱해줌으로써 상기 수학식 13에 비해서 미래의 신호대잡음비와 현재의 신호대잡음비의 차이를 더 크게 하여 보상을 설정할 수 있다.

[0085]

다음으로 선택부(320)는 상기 설정된 자율적인 행동 학습(Q-Learning)의 상태, 행동에서 임의의 상태값 및 상태값에서 가능한 하나 이상의 행동값을 선택한다. 여기서 상태 및 행동은 정해지지 않은 데 반하여, 상태값 및 행동값은 구체적으로 정해진 값을 의미한다. 예를 들어 상태값 3 및 행동값 a_{35} 가 각각 의미하는 것은 3개의 선택 릴레이 상태와 선택 릴레이 개수가 3개에서 5개로 변경되는 행동이다.

[0086]

다음으로 연산부(330)는 상기 선택된 상태값 및 행동값을 기초로 이에 대응하는 하나 이상의 보상값을 수학식 13에 의거하여 연산한다.

[0087]

수학식 12를 참조하면, 자율적인 행동 학습(Q-Learning)에 의한 학습을 위해 보상값은 두 종류의 피드백으로 구성된다. 하나는 현재 취한 행동으로 인한 즉각적인 보상값 $r(s, a)$ 이고, 다른 하나는 현재 행동으로 인해서 나타나는 미래 환경에 대한 보상값인 Q-Value $Q(s', a')$ 이다.

[0088]

그 후에 연산부(330)는 반복 연산을 통하여 Q-Value들로 구성된 Q-Table을 업데이트 한다. 이 때 기본적인 Q-Function은 앞에서 설명한 수학식 12와 같다.

[0089]

마지막으로 제어부(340)는 상기 하나 이상의 보상값 중에서 최고의 보상값을 가지는 행동의 릴레이 개수를 최적의 릴레이 개수로 결정한다.

[0090]

이하에서는 자율적인 행동 학습(Q-Learning)을 이용한 협력 릴레이 선택 방법의 실시예에 대해 보다 상세히 설명한다.

[0091]

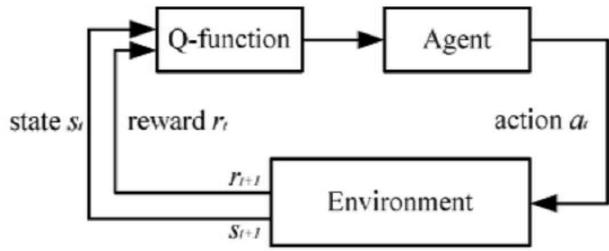
본 발명의 실시예에서 협력통신에 모든 릴레이들이 참여하고, 모든 릴레이들의 인덱스 집합은 수학식 9와 같이

정렬된다고 가정한다.

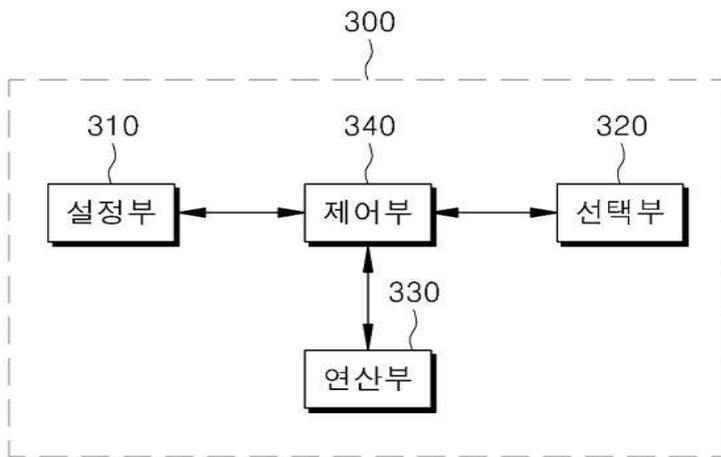
- [0092] 설정부(310)는 "Q-Learning 상태"를 협력통신에서 선택된 협력 릴레이 개수로, "Q-Learning 행동"을 상기 릴레이 개수의 변화로 설정하고 "Q-Learning 보상"을 수학적 식 13으로 정의되는 신호대잡음비에 관한 함수값으로 각각 설정한다(S410).
- [0093] 다음으로 선택부(320)는 상태값 또는 행동값을 선택하기 전에 Q-Table을 초기화하고, 초기 상태값을 랜덤하게 선택한다. 또한 선택부(320)는 현재 상태에서 가능한 모든 행동값들로부터 하나의 행동값을 선택한다(S420).
- [0094] 전체 릴레이 개수가 5인 경우를 예로 들면, 선택부(320)는 초기 상태값을 1 내지 5 중에서 랜덤하게 선택할 수 있다. 선택부(320)가 현재 상태값을 3으로 선택한 경우, 현재의 선택 릴레이 수는 3개를 의미한다. 3의 상태값에서 가능한 행동값은 5개(a_{31} , a_{32} , a_{33} , a_{34} , a_{35})이다. 즉 행동값 a_{31} 은 현재 선택 릴레이 개수를 3개에서 1개로 변경시키는 것을 나타내고, 행동값 a_{35} 는 현재 선택 릴레이 개수를 3개에서 5개로 변경시키는 것을 나타낸다.
- [0095] 선택부(320)는 5개의 행동값 중에서 하나를 선택하여 현재 3개인 선택 릴레이에서 1개 내지 5개의 릴레이로 변경할 수 있다. 다만 3은 현재의 릴레이 개수와 같기 때문에 무의미한 행동값으로 볼 수 있다. 이와 같이 선택부(320)는 5개의 모든 행동값을 선택할 수 있는데 이는 변화된 선택 릴레이 수가 현재의 릴레이 수에 의해 제한받지 않기 때문이다.
- [0096] 릴레이 선택 장치(300)는 선택부(320)가 선택한 행동값을 이용하여 미래 상태 s' 로 이동한다. 이 경우 제어부(340)는 가능한 모든 행동값들에 기반하여, 미래 상태 s' 가운데 Q-Value의 최대값을 가지는 행동값 a' 을 결정한다.
- [0097] 예를 들어서 선택부(320)가 현재 상태값 3에서 행동값으로 4를 선택한 경우 수학적 식 12에 의해, Q-Value는 $Q(3,4)=r(3,4)+ \gamma \max Q(s', a')$ 이다. 즉시 보상값인 $r(3,4)$ 은 수학적 식 13으로 정의된다. 여기서 $\gamma(0<\gamma<1)$ 는 할인 계수를 의미하는데, 할인 계수가 0.5인 경우에, $Q(3,4)=r(3,4)+ 0.5\max Q(s', a')$ 이다. 여기서 $Q(s', a')$ 는 미래 상태(s')의 Q-Value이며 선택부(320)는 $Q(s', a')$, 구체적으로는 $Q(4,1)$, $Q(4,2)$ $Q(4,3)$, $Q(4,4)$ 및 $Q(4,5)$ 중에서 최대(max)의 $Q(s', a')$ 을 가지는 행동값 a' 를 결정한다.
- [0098] 상기의 방법에 의해 연산부(330)는 상기 수학적 식 12를 이용해서 계속해서 Q-Value를 업데이트 한다(S430). 그리고 제어부(340)는 업데이트된 Q-Value로 Q-Table을 생성한다.
- [0099] 본 발명의 실시예에서, 현재 상태(s)와 미래 상태(s')는 상대적인 개념에 해당한다. 즉 연산부(330)가 현재 상태(s)에서의 행동(a)에 의한 Q-Value를 연산한 후에, 상기 현재 상태는 과거가 되고, 상기 미래 상태는 현재 상태가 될 수 있다. 이에 따라 제어부는 상기 미래 상태(s')를 현재 상태(s)로 설정한다.
- [0100] 이러한 과정을 통해 제어부(340)는 업데이트 되는 Q-Value로 구성되는 Q-Table에서 최대의 Q-Value를 가지는 협력 릴레이 개수를 결정한다(S440). 예를 들어 $Q(3,5)$ 의 Q-Value 값이 최대인 경우($a=5$), 바꿔 말하면 릴레이 수가 5개로 변하는 경우, 제어부(340)는 5개를 협력 릴레이 개수로 결정한다.
- [0101] 위와 같은 릴레이 선택 방법의 각 단계에서, 본 발명의 실시예에 따른 릴레이 선택 장치는 단순히 Q-Learning을 위한 행동값을 취하고 그 행동값에 대한 보상값을 받은 후 Q-Table을 업데이트 하게 된다.
- [0102] 한편 본 발명의 실시예에 따른 릴레이 선택 방법은 릴레이의 개수를 선택 결정하는 것으로 설명하였으나, 선택되는 릴레이의 조합을 결정하는 것에도 적용할 수 있다. 예를 들어, 5개의 릴레이($R1$, $R2$, $R3$, $R4$, $R5$) 중에서 최적의 릴레이의 조합을 결정하는 과정에도 적용할 수 있다.
- [0103] 또한, 본 발명의 실시예에 따른 릴레이 선택 방법에서 보상값은 상기 수학적 식 8을 두 번만 계산하면 되므로 알고리즘적인 복잡도와 계산량이 크게 감소하는 반면, 이하의 모의실험 결과를 통해 알 수 있듯이 수학적 분석을 통한 릴레이 선택 기법과 비교했을 때 본 발명은 적은 개수의 릴레이를 사용하지만 거의 유사한 SER 성능을 얻는 것을 보인다.
- [0104] 도 5 내지 도 10은 본 발명의 실시예에 따른 Q-Learning을 이용한 릴레이 선택 방법에 대한 실험 결과를 나타낸 그래프이다. 도 5 내지 도 10의 인덱스 중에서, "Scheme in [4]"는 상기 설명한 MRC에 기반한 릴레이 기법을 나타내며, "Proposed"는 본 발명을 나타낸다.
- [0105] 도 5 내지 도 10을 참조하면, 본 발명에 따른 Q-Learning을 이용한 협력 통신을 위한 릴레이 선택 장치 및 방법의 실험예를 통하여 본 발명의 효과를 검증할 수 있다.

- [0106] 본 실험에 있어서, 통신을 위해서 QPSK (Quadri-Phase Shift Keying) 변조 방식을 사용하고 비트 에너지를 1로 정규화한다. 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 가진다고 가정한다.
- [0107] 또한, 잡음 분산은 1로 정규화하였으며, 하나의 전송 블록에서 채널 계수는 변하지 않으나 서로 다른 블록 사이의 채널 계수는 독립적으로 변한다고 가정한다. 협력통신 네트워크에서 중요하게 다루어야 하는 송신 전력의 경우, 비교되는 릴레이 선택 기법과 본 발명의 하나의 실시예에 따른 릴레이 선택 기법의 복잡도를 낮추기 위해서 본 발명의 하나의 실시예에서는 동등 전력 할당을 사용한다.
- [0108] 동등 전력 할당은 전체 전력 P 절반은 소스와 선택되어진 협력 릴레이 집합에 각각 할당되고, 협력 릴레이 집합에 할당된 전체 전력 P 의 절반은 다시 선택된 각 협력 릴레이들에게 동등하게 분배되는 것을 가정한다.
- [0109] 도 5 내지 도 7까지 보이는 실험 결과는 이용 가능한 전체 릴레이 개수가 $N=10$ 임을 가정한 것이고, 도 5부터 도 8까지 보이는 실험 결과는 수학적 13의 보상 방안을 이용한 것이다.
- [0110] 도 5는 모든 릴레이가 참여할 때(All relays($N=10$)), MRC에 기반한 릴레이 선택 방법이 사용되었을 때(Scheme in [4]), 그리고 본 발명의 하나의 실시예에 따른 Q-Learning을 이용한 릴레이 선택 방법이 사용되었을 때(Proposed)의 SER 성능을 보여준다.
- [0111] 도 5에서 $SNR = 8$ dB까지 본 발명의 하나의 실시예에 따른 방법(Proposed)과 MRC에 기반한 릴레이 선택 방법(Scheme in [4])의 SER 성능이 매우 유사하고, 그 이후에는 제안 기법의 SER 성능이 약간 악화되는 것을 관찰할 수 있다.
- [0112] 도 6은 협력통신을 위해서 사용된 평균 릴레이 개수를 보여준다. 여기서, 본 발명의 실시예에 따른 방법(Proposed)이 종래 기술에 따른 방법(Scheme in [4]) 보다 약 2.5개의 릴레이를 절약하고, 모든 릴레이가 협력 통신에 참여했을 때(All relays($N=10$)) 보다 약 7.5개의 릴레이를 더 적게 사용하는 것을 확인할 수 있다.
- [0113] 도 7은 $SNR = 10$ dB에서 이용 가능한 릴레이 개수가 증가함에 따라서 협력통신을 위해 선택된 평균 릴레이 개수를 보여준다. 이 결과로부터, MRC에 기반한 방법(Scheme in [4])과 다르게 본 발명의 하나의 실시예에 따른 방법(Q-Learnin)은 이용 가능한 릴레이 개수가 증가함에도 불구하고 평균적으로 약 2.5개의 고정된 개수의 릴레이를 선택하는 것을 보여준다.
- [0114] 도 8은 협력통신에 참여하는 다양한 릴레이 개수에 따른 SER 성능을 보여준다. 도 7 및 도 8에서 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 SNR 이 8 dB 보다 작을 때, 협력 릴레이 개수가 약간 감소하는 것은 SER 성능에 거의 영향을 미치지 못하고, 더 나아가서 SNR 이 8 dB 이상일 때는 비록 SER 성능이 약간 저하되지만 협력 릴레이 개수를 급격하게 줄일 수 있다는 것을 보여준다.
- [0115] 도 9 및 도 10은 상기 수학적 14의 보상 방안을 이용했을 때에 대한 실험 결과를 보여준다. 앞서 언급한 바와 같이, 수학적 14는 수학적 13에 $state_{next}$ 항을 추가함으로써 보상값의 격차를 크게 만든 것이다. 도 7 및 도 8에서 볼 수 있듯이, 수학적 14의 보상을 이용한 릴레이 선택 방법은 종래 기술에 따른 방법과 비교했을 때 SER 성능과 협력 릴레이 개수에 대한 측면에서 매우 비슷한 성능을 보인다. 하지만 수학적 13을 사용할 때와 비교해서 보상의 정의만 바뀌었으므로, 여전히 수학적 분석을 통해서 릴레이를 선택하는 MRC 기반의 선택 방법(Scheme in [4])에 비해서 알고리즘적인 복잡도와 계산량은 적다. 이러한 실험 결과는 Q-Learning에서 보상 r 을 정의하는 형태에 따라서 시스템의 목적을 다르게 설정할 수 있음을 보여준다.
- [0116] 상기와 같은 실험예들은 본 발명의 하나의 실시예에 따른 릴레이 선택 방법이 행동에 대한 보상값을 주고 Q-Table을 업데이트하는 간단한 릴레이 선택 과정과 함께 협력 릴레이 개수를 급격하게 감소시키지만 우수한 SER 성능을 얻을 수 있음을 보여준다.
- [0117] 본 발명의 실시예에 따른 자율적인 행동 학습(Q-Learning)을 이용한 릴레이 선택 방법은 자율적인 행동 학습(Q-Learning)을 위하여 상태, 행동, 보상을 적절히 정의하는 것에 의해서 복잡한 수학적 계산 없이 우수한 SER 성능을 얻는 시스템이 도출될 수 있다.
- [0118] 특히, 종래 기술에 따른 방법에서는 이용 가능한 릴레이 개수가 증가함에 따라서 협력 릴레이 개수 또한 선형적으로 증가하였지만, 본 발명의 하나의 실시예에 따른 Q-Learning을 이용한 릴레이 선택 방법은 평균적으로 약 2.5개의 협력 릴레이만을 사용하였다.

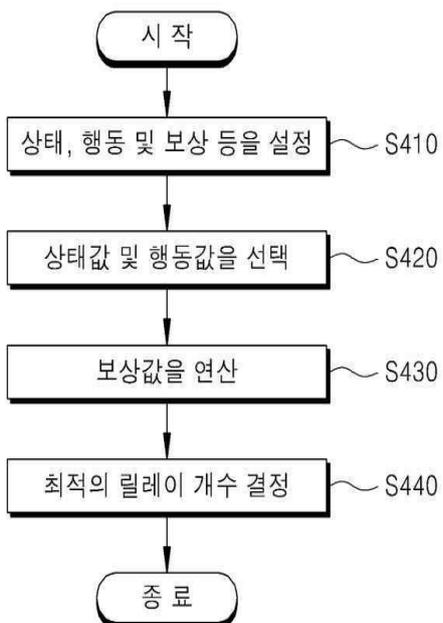
도면2



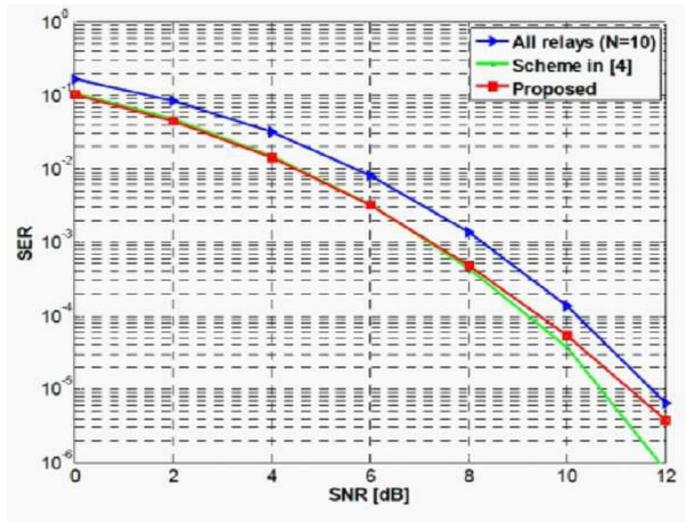
도면3



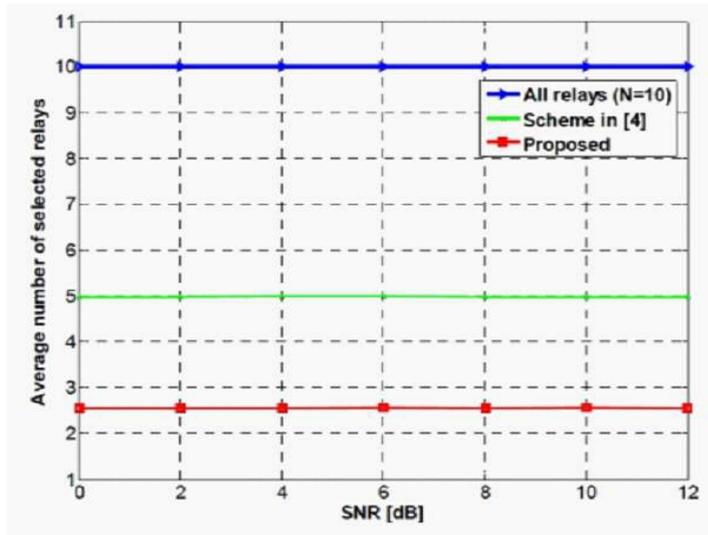
도면4



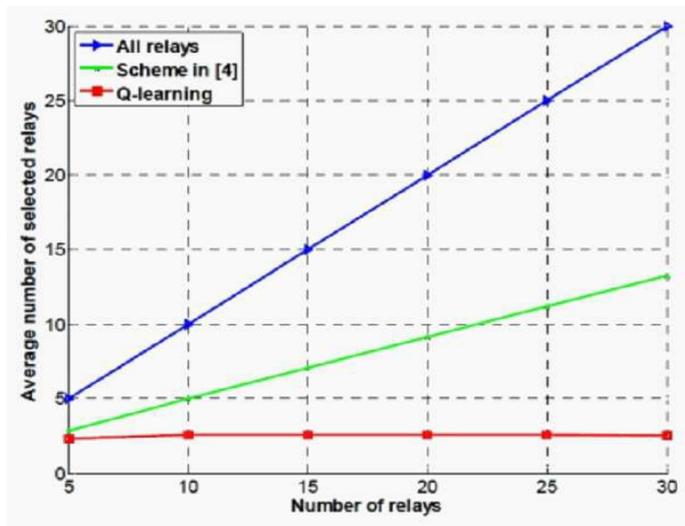
도면5



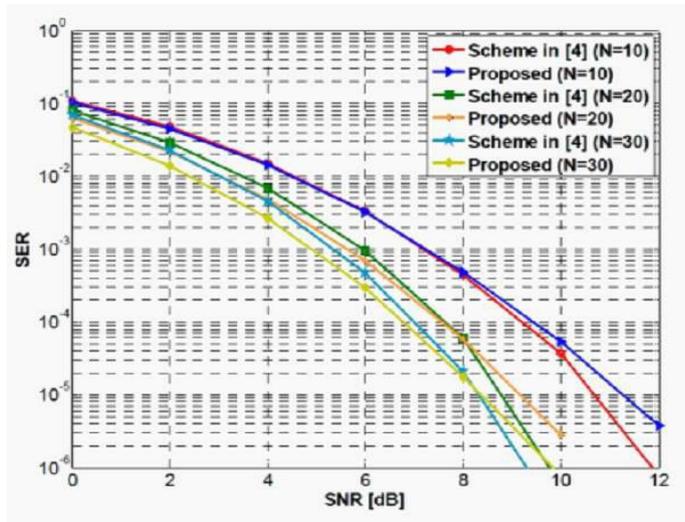
도면6



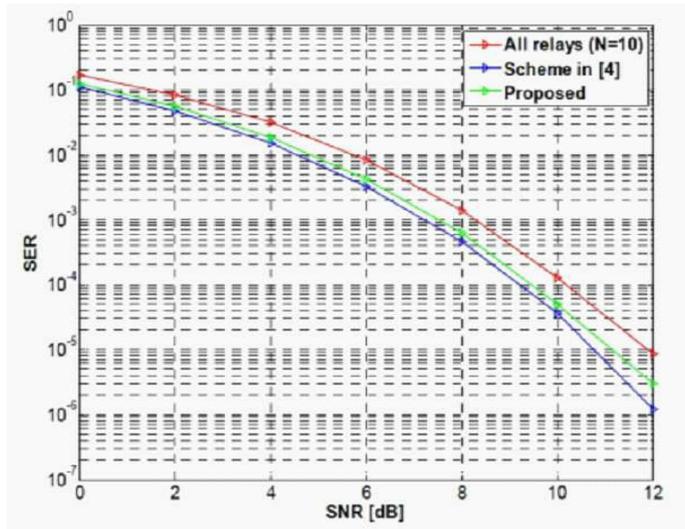
도면7



도면8



도면9



도면10

