무선 인지 시스템을 위한 Q-learning 기반 채널접근기법

A Q-learning based channel access scheme for cognitive radios

이 영 두* 구 인 수**
Youngdoo Lee Insoo Koo

요 약

가용 주파수 고갈 문제를 해결하기 위하여 제안된 무선인지기술은 특정 주파수 대역에 대해 사용면허를 가진 주사용자가 사용하지 않는 유휴채널에 접근하여 통신을 수행함으로써 주파수 효율을 향상시키는 차세대 통신기술이다. 주사용자의 유휴채널을 사용하기 위해서는 해당 채널을 현재 주사용자가 점유하고 있는지를 정확히 판단하여야 한다. 분산형 무선인지 네트워크에서 독립적으로 채널을 센싱하는 무선인지 기기의 경우 센싱의 결과가 노이즈, 쉐도윙, 페이딩과 같은 채널 환경에 영향을 많이 받으며 심지어 주사용자가 요구하는 간섭률을 보장하지 못하는 결과를 초래한다. 따라서 본 논문에서는 주사용자가 요구하는 최소 간섭량을 보장하는 동시에 기회주의적으로 채널에 접근하여 인지시스템의 처리율(처리율)을 향상시키는 Q-learning 기반의 채널접근기법을 제안한다. 제안하는 기법은 사전 학습 단계에서 주사용자의 채널사용 패턴을 Q-learning으로 학습하고 이를 Q-learning 기반 채널접근 단계에서 실제로 적용함으로써 스펙트럼 센싱 성능을 향상시킨다. 모의실험을 통해 AWGN 및 레일레이 페이딩 무선 환경에서 주사용자에 대한 간섭량 및 처리율 성능이 기존의 에너지 검출 방법에 비해 더 우수함을 확인하였다.

ABSTRACT

In distributed cognitive radio networks, cognitive radio devices which perform the channel sensing individually, are seriously affected by radio channel environments such as noise, shadowing and fading such that they can not properly satisfy the maximum allowable interference level to the primary user. In the paper, we propose a Q-learning based channel access scheme for cognitive radios so as to satisfy the maximum allowable interference level to the primary user as well as to improve the throughput of cognitive radio by opportunistically accessing on the idle channels. In the proposed scheme, the pattern of channel usage of the primary user will be learned through Q-learning during the pre-play learning step, and then the learned channel usage pattern will be utilized for improving the sensing performance during the Q-learning normal operation step. Through the simulation, it is shown that the proposed scheme can provide better performance than the conventional energy detector in terms of the interference level to primary user and the throughput of cognitive radio under both AWGN and Rayleigh fading channels.

☞ keyword : cognitive radio(무선인지), channel access(채널 접속), spectrum sensing(스펙트럼 센싱), Q-learning(Q-learning)

1. 서 론

무선 서비스의 고속화 및 멀티미디어화에 대한 요구가 급속도로 증가함에 따라 다양한 형태의 통신기기 및 표준들이 개발/논의되고 있으며, 이러한 양상은 한정적 자원인 주파수에 대한 부족현상으로 이어져 기존의 고정 주파수 할당 정책의 한계로서 심각한 문제로 대두되었다. 이에 대한 해결책으로 정부가 주도적으로 관리하던 기존 주파수 정책에서 주파수의 용도 및 기술조건 등에 관한 규제를 완화하고 ISM 밴드와 같은 비허가 대역의 확대를 추진하는 등의 개방형 주파수 정책이 추진되고 있다.

주파수 사용률에 대한 FCC(Federal Communications Commission, 미국연방통신위원회)의 연구조사에 따르면 시간적으로나 지역적으로 평균 주파수 사용률은 약 15%~85% 정도로 크게 변화고 있

* 정 회 원 : 울산대학교 전기전자정보시스템공학부
박사과정 leeyd1004@naver.com
** 정 회 원 : 울산대학교 전기전자정보시스템공학부 부교수
iskoo@ulsan.ac.kr(교신저자)
[2011/01/21 투고 - 2011/01/26 심사 - 2011/04/07 심사완료]

으며 이는 효율적인 주파수 사용이 제대로 이루어지고 있지 않음을 나타낸다. 주파수 사용률에 대한 관심은 주파수 공유 기술 및 무선 환경에 따라 지능적으로 변조방식, 출력 등을 제어하여 최적의 상태로 통신을 수행하는 무선통신 기술에 대한 연구를 크게 일으켰고, 대표적인 기술로 무선인지(Cognitive Radio, 이하 CR) 기술이 연구되고 있다[1-3].

CR 기술의 주요한 동작은 다음과 같다. 먼저 사용하고자 하는 주파수 대역에 대해 스펙트럼 센싱(spectrum sensing)을 수행하여 해당 주파수를 고정적으로 사용하는 주사용자(Primary User, 이하 PU)의 주파수 사용 유무를 확인하고, 스펙트럼 센싱의 결과를 기반으로 PU가 감지되지 않는 주파수 대역을 선택하는 스펙트럼 결정(spectrum decision)을 수행한다. 그리고 결정된 주파수 대역의 무선 환경을 고려하여 최적의 전송 파라미터를 찾는 동시에 다른 CR 기기간의 충돌을 피하기 위한 스펙트럼 공유(spectrum sharing) 단계를 거쳐 통신을 수행한다. 만약 스펙트럼 결정하여 사용하고 있는 주파수 대역에 PU가 통신을 수행하게 되면 다른 유휴 주파수를 찾아 이동하는 스펙트럼 이동(spectrum mobility)을 수행한다[4].

상기 주요 기능 가운데 스펙트럼 센싱은 CR 기기가 야기 할 수 있는 PU에 대한 간섭을 제거하고 사용 가능한 유휴 주파수 자원을 검출 할 수 있도록 하는 핵심적인 CR의 요소 기술로서 대표적인 방법들로는 에너지 검출(energy detection), 정합필터 검출(matched filter detection), 신호 특징 검출(feature detection) 등이 있다. 이중 에너지 검출은 사용하고자 하는 주파수 대역에서 측정한 신호의 에너지 레벨을 확인하고 일정한 임계값 이상일 경우 PU가 해당 주파수 대역에 존재하는 것으로 판단하는 방법으로 다른 방법들에 비해 검출 성능은 떨어지나 낮은 복잡도와 구현 비용 등의 장점으로 인해 많이 사용되고 있다[4-6].

그러나 에너지 검출은 그 복잡도와 구현의 간편성에 비해 그 특성상 단지 주파수 대역에 PU의 존재 유무만을 판단할 수 있으므로 그 외의 주변 무선 환경에 대해서는 파악하는 것이 불가능하며, 각 CR 기기 별로 독립적으로 스펙트럼 센싱이 수행될 때 주변 장애물에 의해 발생하는 쉐도윙(shadowing) 및 다중 경로 페이딩(multi-path fading) 등으로 인하여 해당 주파수 대역을 PU가 사용하고 있음에도 불구하고 이를 검출하지 못하는 경우가 발생하게 된다.

따라서 본 논문에서는 에너지 검출을 사용하며 각 CR 기기에 의해 독립적으로 수행되는 스펙트럼 센싱의 성능을 향상시키기 위하여 Q-learning 기반의 채널접근 기법을 제안한다. 제안하는 기법에서 각 CR 기기는 자체적인 스펙트럼 센싱후 본 논문에서 제안하는 MAC 프로토콜을 기반으로 서로 데이터를 주고받으며, Q-learning의 초기 학습(learning)을 위해 일정한 시간 동안에는 에너지 검출기의 결과를 그대로 사용하고 그 이후에 Q-learning 기반의 스펙트럼 결정을 수행하여 채널에 접근한다. 성능 평가 요소로서 PU가 채널을 사용하고 있는 동안 CR 기기가 채널에 접근함으로써 PU의 통신이 불능상태가 되는 간섭에 대한 요구 간섭률과 CR 기기의 처리율(throughput) 성능을 고려한다.

본 논문의 구성은 다음과 같다. 2장에서 본 논문에서 고려하는 시스템 모델을 간략히 기술하고, 3장에서 제안하는 미니 슈퍼프레임 기반의 MAC 프로토콜을, 4장에서는 제안하는 채널접근 기법을 설명한다. 5장에서 AWGN 및 레일레이(Rayleigh) 페이딩 채널 하에서 앞서 고려한 성능 평가 요소에 대한 모의실험 결과를 도시한다. 그리고 6장에서 결론을 맺는다.

2. 시스템 모델

본 논문은 분산형 CR 네트워크에서 이동형의 CR 기기들이 하나의 채널에 접근하여 서로 데이터를 송수신함으로써 원격지의 AP(access point)로 데이터를 전송하며, CR 기기들은 센싱 정보를

공유하는 것 없이 개별적인 스펙트럼 결정을 수행하는 것으로 가정한다. 그리고 스펙트럼 결정은 다수의 채널 중 사용 가능한 유휴 채널을 선택하는 문제가 아닌 하나의 채널에 대한 접근시간(access timing)을 결정하는 것으로 재 정의한다.

본 장에서는 앞서 기술한 CR 네트워크 기반 하에서 이동형 CR 기기가 사용하고자 하는 채널의 PU 사용유무를 판단하기 위해 어떻게 스펙트럼 센싱을 수행하는지를 전반적으로 설명하고, 이어 다음 장에서 각 CR 기기가 채널에 접근하기 위해 사용하는 본 논문에서 제안하는 MAC 프로토콜을 설명한다.

일반적으로 CR 기기는 유휴 주파수 대역을 검색하거나 사용하고자 하는 주파수 대역에 PU가 활동 중인지를 주기적으로 확인하는 과정을 수행하는데 이것을 스펙트럼 센싱이라 한다. 본 논문에서는 에너지 검출기(energy detector, 이하 ED)를 사용하여 해당 주파수 대역을 센싱한다. ED는 이진 가설검정(binary hypothesis test)을 사용하여 PU의 신호를 검출하며, 고려되는 신호의 형태는 아래와 같다.[4-6]

$$r(t)=\begin{cases} n(t) & : H_0 \\ s(t)+n(t) & : H_1 \end{cases} \quad (1)$$

$r(t)$는 CR 기기가 수신하는 신호를 의미하며, $n(t) \sim CN(0,\sigma_n^2)$는 복소 백색 가우시안 노이즈(complex AWGN)를, $s(t)$는 PU의 신호를 나타낸다. H_0는 PU의 신호가 검출되지 않아 해당 주파수 대역이 유휴 상태임을 의미하며, H_1은 PU에 의해 해당 주파수 대역이 사용되고 있음을 의미한다.

수식(2)과 같이, 수신된 신호에 대해 M 개의 샘플들을 수집하여 구한 평균값 r_{avr}는 하나의 가설 통계량(test statistic)으로 사용될 수 있으며, M > 30 일 때 r_{avr}는 중앙극한정리(central limit theory)에 의해 하나의 가우시안 랜덤변수(random variable)로 고려할 수 있다. 따라서 H_0과 H_1은 수식(3)의 값들을 가지는 가우시안 분포로 근사화될 수 있다.

$$r_{avr}=\sum_{j=1}^{M}|r_j|^2 \quad (2)$$

$$\begin{cases} \mu_0 = M\sigma_n^2,\ \sigma_0^2 = 2M\sigma_n^4 & : H_0 \\ \mu_1 = M(\gamma+\sigma_n^2),\ \sigma_1^2 = 2M\sigma_n^2(2\gamma+\sigma_n^2) & : H_1 \end{cases} \quad (3)$$

수식(3)에서 $\gamma = |s^2|/M$는 PU 신호의 평균전력이며, M=2*T*W 로서 센싱시간 T 와 센싱하는 주파수의 대역폭 W의 곱으로 결정된다.

r_{avr}에 의한 스펙트럼 센싱의 결과는 수식(4)에 의해 도출되며, 임계값 λ는 PU의 통신을 보호하기 위하여 요구된 오검출(probability of miss detection, P_m) 값에 기반하여 수식(5)과 (6)을 통해 유도된다[6].

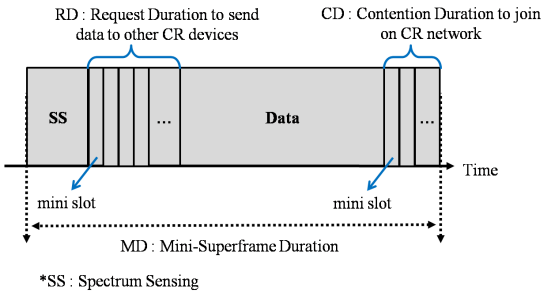
$$r_{avr} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \quad (4)$$

$$\begin{aligned} P_m &= 1-Pd = 1-\Pr(r_{avr}>\lambda|H_1) \\ &= 1-Q\left(\frac{\lambda - M\sigma_n^2 - M\gamma}{\sigma_n\sqrt{2M\sigma_n^2+4M\gamma}}\right) \end{aligned} \quad (5)$$

$$\lambda = Q^{-1}(1-P_m)\left(\sigma_n\sqrt{2M\sigma_n^2+4M\gamma}\right) + M\sigma_n^2 + M\gamma \quad (6)$$

수식(5)와 (6)의 Q-function은 아래와 같이 정의된다.

$$Q(x)=\frac{1}{\sqrt{2\pi}}\int_x^{+\infty} e^{-\tau^2/2}d\tau \quad (7)$$



(그림 1) 미니 슈퍼프레임의 구조

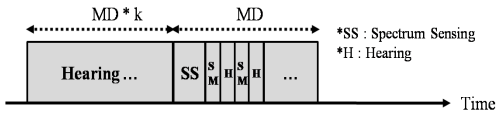
3. 제안하는 MAC 프로토콜

앞서 설명한 바와 같이 CR 기기들은 자신의 정보를 AP까지 보내기 위해서 먼저 사용하고자 하는 채널에서 데이터를 송신하기 위한 채널접근기회를 찾아야 한다. 이를 위해 해당 채널을 사용하고 있는 PU와 다른 CR 기기들을 고려한 채널접근을 시도하여야 한다.

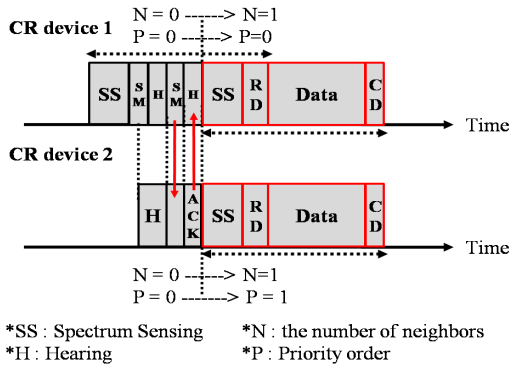
본 논문에서는 (그림 1)의 미니 슈퍼프레임 구조를 제안함으로써 CR 기기간의 동시 채널 접근으로 인한 충돌을 피하고 PU를 감지하는 동작을 수행하도록 한다. 또한 제안하는 MAC 프로토콜은 다음 장에서 기술될 Q-learning의 보상함수의 보상값을 계산하는 기준정보를 제공하는 역할을 수행한다.

하나의 미니 슈퍼프레임은 스펙트럼 센싱을 수행하기 위한 SS(spectrum sensing) 구간, CR 기기간 RTS/CTS 패킷을 주고받기 위한 RD(request duration), 실제 데이터를 송수신하는 Data 구간, 그리고 새로운 CR 기기의 네트워크 참여를 위한 CD(contention duration)로 구성되며, RD와 CD는 각각 다수의 미니 슬롯(mini slot)으로 구성된다.

CR 기기가 통신을 위하여 임의의 채널을 선택하면, 해당 CR 기기는 CR 네트워크에 자신을 등록시키기 위해 미니 슈퍼프레임이 수신되기를 (그림 2)와 같이 k 개의 MD(mini-superframe duration) 동안 기다린다. 만약 k 개의 MD 기간동안 미니



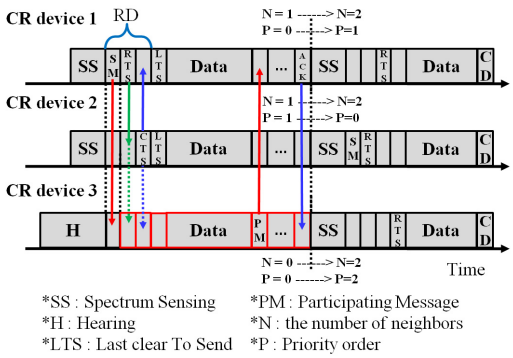
(그림 2) 미니 슈퍼프레임의 동기화(1)



(그림 3) 미니 슈퍼프레임의 동기화 과정(2)

슈퍼프레임을 수신하지 못하면 그림과 같이 독자적으로 미니 슈퍼프레임을 구성하여 송신하며 다른 CR 기기가 나타나기를 기다린다. 이때 CR 기기는 SS 수행 후 만약 채널이 PU에 의해 사용되지 않는 것으로 판단되면 미니 슈퍼프레임의 동기 신호인 SM(synch message) 패킷과 이에 대한 다른 CR 기기의 응답인 ACK 패킷을 기다리는 동작을 주기적으로 수행한다.

(그림 3)은 채널에 다른 CR 기기가 출현한 경우에 미니 슈퍼프레임에 대한 동기화를 보여준다. CR device 1은 주기적으로 SM 패킷 송신과 ACK 패킷 수신 대기를 수행 중에 CR device 2로부터 ACK 패킷을 수신하면 해당 채널에 다른 CR 기기가 출현했음을 인지하게 되어 SM 패킷 송신을 멈추고 ACK 패킷을 기반으로 일반적인 미니 슈퍼프레임 구조에 기반하여 동작한다. 그리고 CR device 2는 SM 패킷을 이용하여 ACK 패킷을 송신한 후 미니 슈퍼프레임 동기화를 수행한다. 이때 CR device 1과 2는 전체 이웃 CR 기기의 수를 나타내는 N 값과 Data 송수신을 위한



(그림 4) 미니 슈퍼프레임의 데이터 전송과정

우선순위(Priority)를 결정하는 P 값을 (그림 3)과 같이 각각 갱신시킨다. 양쪽 모두 N 값이 동일하게 증가하지만 먼저 채널을 사용하고 있던 CR device 1은 P=0을 그대로 유지하고, CR device 2는 이후에 왔기 때문에 P=1 값을 가진다.

(그림 4)는 추가적인 CR 기기의 등장과 동시에 각 CR 기기간 데이터를 송수신하는 과정을 보여준다. 미니 슈퍼프레임은 새롭게 채널에 접근하는 CR 기기의 동기화를 위해 RD의 첫 미니 슬롯에는 SM 패킷을 송신하도록 되어 있으며 다음 미니 슬롯부터는 P 값을 기반으로 미니 슬롯을 통해 RTS 패킷을 송신한다. 초기 RD에 대한 미니 슬롯 사용 우선순위는 P 값을 기반으로하지만 이후에는 CR 기기간의 공평성(fairness)을 위하여 라운드 로빈(Round robin) 방식으로 스케줄링된다. 그리고 RD의 미니 슬롯 사용에 대해 최우선순위를 지니는 CR 기기는 앞서 설명한 바와 같이 RD의 첫 미니 슬롯에 미니 슈퍼프레임에 동참하고자 하는 CR 기기들의 동기화를 위해 SM 패킷 송신의 의무가 주어진다. (그림 4)에서 CR device 1은 현재 RD의 미니 슬롯 사용에 대해 최우선순위를 가지므로 SM을 송신한 후, 만약 다른 CR 기기, 곧 CR device 2로 전송할 데이터가 있다면 RTS 패킷을 송신하고 다음 미니 슬롯에서 CR device 2로부터 CTS 패킷을 수신 받게 된다. 만약 없다면 침묵하고 CR device 2로 RTS 패킷 송신에 대한 기회가 돌아가게 된다. LTS(Last clear To Send)는 최하위 우선순위를 가진 CR 기기의 RTS 패킷에 대한 응답 CTS 패킷 전송을 위한 미니 슬롯이다. 결과적으로, RD를 통해 오직 한 쌍의 CR 기기들만이 Data 구간을 사용하게 된다. 추가로, 일단 RTS/CTS 패킷에 의해 CR 네트워크 상에 전송 링크 설정에 대한 알림이 끝이 난후 RD에 미니 슬롯이 남아 있으면 그것들은 Data 구간으로 편성되어 Data 전송에 사용되고, 항상 데이터를 수신하는 CR 기기는 Data 구간 끝에서 수신에 대한 ACK 패킷을 전송한다. 미니 슈퍼프레임에 동참하고자 하는 CR device 3은 CR device 1이 송신한 SM 패킷을 수신함으로써 현재 운영되고 있는 미니 슈퍼프레임의 CD의 위치정보를 얻게 되고 CD가 시작되는 시간까지 대기한다. 그리고 CD의 미니 슬롯 상에서 CSMA/CA 방식으로 자신의 PM(participating message)을 전송하고 CD의 마지막 미니 슬롯에서 RD의 미니 슬롯 사용에 대한 최우선순위를 가지는 CR device 1으로 부터 ACK를 전송받아 미니 슈퍼프레임을 동기화함으로써 해당 채널의 CR 네트워크에 참여한다. 이때 전송기회에 대한 공평성을 위하여 CR device 1의 P=1로, CR device 2의 P=0으로, 그리고 새롭게 참여한 CR device 3의 P=2로 조정되고 이후로 계속해서 라운드 로빈 방식으로 스케줄링이 진행된다.

4. 제안하는 채널접근기법

일반적으로 ED는 채널에서 수집된 신호의 에너지 레벨만을 고려함으로 극심한 페이딩에 의한 열악한 채널상태에서는 PU 신호를 제대로 검출하기 어렵다. 따라서 본 논문에서는 AWGN 및 레일레이(Rayleigh) 페이딩 하에서 PU에 대해 요구 간섭률(Interference rate to PU)이 주어질 때 이를 준수하면서 채널접근기회를 찾아 데이터 송수신을 수행 할 수 있는 Q-learning 기반 채널접근기법을 제안한다.

4.1 Q-learning

Q-learning 알고리즘은 대표적인 강화학습(reinforcement learning) 알고리즘 중 하나로써 주변의 상태(state, s)뿐 아니라 그에 상응하는 행동(action, a)을 하나의 순서쌍으로 함께 고려한 Q(s,a) 값에 기반하여 최적의 정책(policy)을 찾아내는 알고리즘이다[7-9].

학습을 수행하는 객체를 에이전트(agent)라하고, 에이전트에 의해 센싱 가능한 상태 공간을 S, 상태 공간 S에서 에이전트의 실행 가능한 행동들의 집합을 A라 할 때, 임의의 상태 s∈S 에 대한 행동 a∈A의 쌍 (s,a)는 다음과 같은 가치 함수(value function)로서 평가될 수 있다.

$$V^{\pi}(s,a)= E[\sum_{t=0}^{\infty}\gamma^{t} r_t(s,a)|s_0 = s, a_0 = a] \quad (8)$$

$\gamma \in$ [0,1) 은 할인율(discount rate)을 의미하며, 가치 함수가 수렴하도록 한다. $r_t(s,a)$ 는 보상함수로서 학습의 결과로부터 얻어지는 강화 값을 의미한다. 가치 함수의 π는 하나의 정책을 표시하는 것이며, 정책은 어떻게 상태 s에 대해 a를 취할 것인지를 나타내는 함수이다.

$$\pi(s) = a,\ s\in S,\ a\in A \quad (9)$$

따라서 최적의 정책은 가치 함수를 최대화 시키는 π가 된다.

$$V^{*} = \max_{\pi} V^{\pi}(s,a) \quad (10)$$

Q-learning은 수식(10)의 최적의 정책을 근사적으로 유도하기 위하여 현재 상태 s_t에서 임의의 행동 a_t를 수행하였을 때 받은 보상값에 대한 결과를 Q(s_t,a_t)에 저장하고 다음 상태 s_{t+1}에 대해 최대의 Q(s_{t+1}, a_{t+1}) 값을 가지는 행동 a_{t+1}을 선택하여 수식(11)과 같이 현재 상태의 Q(s_t,a_t) 값을 갱신한다.

$$Q(s_t,a_t)= (1-\alpha)Q(s_t,a_t) + \alpha\{r_{t+1} + \gamma \max_{a\in A} Q(s_{t+1},a_{t+1})\} \quad (11)$$

$r_{t+1} = r(s_{t+1},a_{t+1})$을 나타내고, $\alpha\in[0,1]$ 는 learning rate로서 학습의 결과로부터 얻어지는 강화 값의 갱신율을 조절하는 역할을 수행한다.

Q-learning에서 최적의 정책을 찾기 위해서는 모든 상태-행동 쌍에 대한 충분한 탐색을 보장해야 한다. 이것은 지역 해(local solution)를 전역 해(global solution)로 오인하는 결과를 방지하기 위해서이다. 이를 위해 exploration-exploitation 방식과 볼쯔만 확률 분포를 따라 행동을 선택하는 방식이 널리 이용되고 있으며 본 논문에서는 exploration-exploitation 방식을 채택한다.

exploration-exploitation 방식에서 exploration은 다음 상태 s_{t+1}를 위한 행동 a_{t+1}을 선택할 때 기존의 Q(s,a)를 고려하지 않고 무작위로 가능한 행동 a_{t+1}을 선택하는 것이며, exploitation은 수식(11)의 $\max_{a\in A} Q(s_{t+1},a_{t+1})$와 같이 최대의 Q(s,$a$) 값을 가지는 행동 a_{t+1}을 선택하는 것이다. exploration과 exploitation은 균등분포 확률변수를 이용하여 각각 선택되는데 ϵ의 확률로는 exploration이 $(1-\epsilon)$의 확률로는 exploitation이 선택된다[10-11].

4.2 제안하는 Q-learning 기반 채널접근기법

본 논문에서 제안하는 기법은 그림 5와 같이 3가지 단계로 구분된다. 초기화(initialization) 단계에서는 Q-learning에 사용되는 파라미터들을 초기화하며, 사전 학습(Pre-play learning) 단계에서는 Q-learning의 초기 학습을 수행한다. 이때 CR 기기는 ED의 스펙트럼 센싱 결과를 고려하여 채널에 PU의 활동 유무를 판단하고 접근한다. Q-learning 기반 채널접근(Q-learning normal operation)

단계에서는 ED가 도출한 센싱 결과에 대해 Q-learning을 수행하고 채널접근 여부를 결정한다.

4.2.1 상태 공간 & 행동집합

일반적으로 ED에 의한 스펙트럼 센싱의 결과는 크게 H_0와 H_1의 2 가지 상태로 나타낼 수 있으므로, 이에 상응하는 상태 공간 $S=\{s_0,s_1|H_0\rightarrow s_0, H_1\rightarrow s_1\}$가 된다. CR 기기는 채널접근의 관점에서 2가지의 행동을 취할 수 있다. 그것은 채널을 "사용한다"와 "사용하지 않는다" 이며, 따라서 $A=\{a_0,a_1|use\rightarrow a_0, unuse\rightarrow a_1\}$로 표현된다.

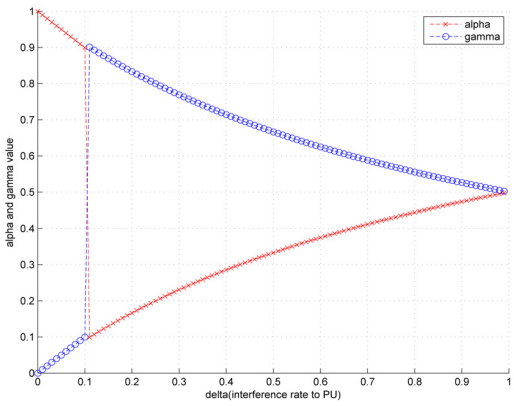
4.2.2 보상함수(reward function)

보상 함수 $r(s,a)$ 는 PU의 채널 사용유무를 나타내는 상태 s_i , $i=\{0,1\}$ 에 대해 CR 기기가 선택한 행동 a_j , $j=\{0,1\}$ 가 결과적으로 CR 기기에 어떤 이익이 되는지를 판단하여 해당 Q(s_i,a_i)의 값을 + 강화 또는 - 강화 시키는 역할을 수행하는 함수이며, (표 1)과 같이 정의된다. (표 1)에서 현재상태(current state)는 ED에 의해 수행된 스펙트럼 센싱의 결과이며, 선택된 행동(selected action)은 Q-learning을 통해 결정된 행동을 의미한다. 그리고 실제 상태(true state)는 CR 기기에 의해 판단되는 PU의 채널 사용유무이며, δ는 현재 CR 기기의 PU에 대한 간섭률을 나타낸다. 여기에서 실제 상태는 PU의 실제 채널 사용유무를 의미하는 것이 아니며 단지 CR 기기의 정의된 MAC 프로토콜을 기반으로 PU의 채널 사용유무를 판단하는 것을 의미한다. 그 판단 기준은 다음과 같다. 정의된 MAC 프로토콜에서 미니 슈퍼프레임을 통해 수신 받아야 할 SM, RTS/CTS, ACK 등과 같은 프로토콜 패킷을 제대로 수신하면 실제 상태를 s_0, 그렇지 않으면 s_1으로 판단한다. 이러한 판단은 PU 신호의 최대전송거리와 전송되는 프레임의 길이가 CR 기기들보다 크거나 같다고 가정될 때 사용가능하다.

보상값은 현재 CR 기기의 PU에 대한 간섭 지수인 δ로 정의되었다. (표 1)에서 쌍 (현재 상태-선택된 행동-실제 상태)이 (S0,a0,S0) 일 때 CR 기기는 Q-learning 기반으로 옳은 판단을 하였으므로 $(1-\delta)$ 를 + 강화 값으로 가지며, (S0,a0,S1)일 때는 오판하였으므로 $-\delta$ 만큼을 - 강화 값으로 가진다. (S0,a1,S0)의 경우에는 오판이므로 $-\delta$ 값을, (S0,a1,S1)의 경우에는 옳은 판단이므로 $(1-\delta)$ 값을 가진다. (S1,x,x)의 경우들은 (S0,x,x)의 경우들과는 반대의 효과를 가진다. (S1,x,x)에서는 ED가 채널이 PU에 의해 사용되고 있다고 판단했으므로 CR 기기가 채널을 사용하지 않아야 하지만 실제로 PU가 사용하고 있지 않다면 도리어 채널접근기회를 상실하는 경우가 된다. 따라서 만약 (S1,a0,S0)과 같이 PU가 채널을 사용하고 있다고 ED가 판단을 내린 상황에서 Q-learning 기반으로 채널접근 유무를 판단하여 채널접근기회를 얻게 되면 위험률 만큼을 보상값으로 주기 위하여 δ 가 + 강화 값으로 주어지고, (S1,a0,S1)의 경우처럼 오판일 때는 ED가 제대로 판단하였는데 이를 번복했으므로 $-(1-\delta)$의 값이 - 강화 값으로 주어진다. (S1,a1,S0)와 (S1,a1,S1) 역시도 동일한 맥락에서 표 1의 값을 가진다.

(표 1) 제안된 보상 함수

Current state (ED decision)	Selected action (Q-learning decision)	True state	Reward
S0	a0(use)	S0	1-δ
		S1	-δ
	a1(unuse)	S0	-δ
		S1	1-δ
S1	a0(use)	S0	δ
		S1	-(1-δ)
	a1(unuse)	S0	-δ
		S1	(1-δ)



(그림 6) $\triangle = 1$ 일 때, δ에 따른 α, γ의 값

4.2.3 적응 학습

본 절에서는 주어진 요구 간섭률의 한도 내에서 처리율 성능을 향상시키기 위해 Q-learning의 할인율 γ와 learning rate α을 가변적으로 조절함으로써 기회주의적으로 채널접근을 시도하는 적응 학습 기법을 제안한다. Q-learning에서 γ는 다음 상태 s_{t+1}와 exploration-exploitation 방식을 기반으로 결정되는 기대 값의 영향력을 조절하는 파라미터이며, α는 수식(11)에서 보는 바와 같이 과거 Q 값과 현재 얻은 전체 강화 값(=보상 값의 합) 사이의 비율을 조절하는 가중치 파라미터이다. α와 γ의 값이 증가 할 때 그 특성상 현재 얻은 전체 강화 값의 비율이 증가하여 처리율 성능을 향상시키지만 또한 이로 인해 간섭률도 증가하게 된다. 그러므로 기회주의적으로 채널접근기회를 얻기 위해서는 주어진 요구 간섭률을 고려하여 α와 γ을 가변적으로 조절하여야 한다.

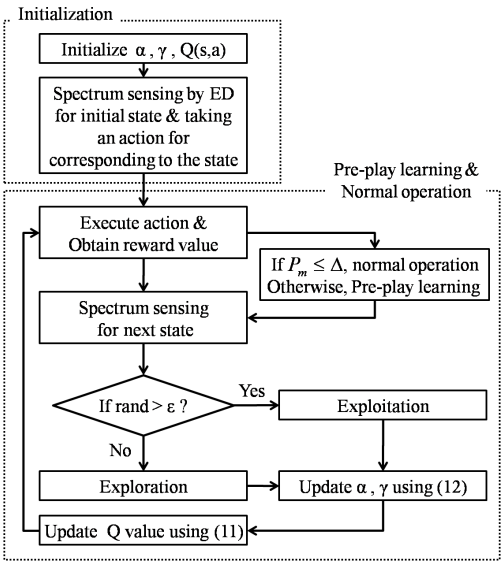
제안하는 적응 학습 기법을 위한 α와 γ는 다음과 같다.

$$\begin{cases} \alpha = -\delta,\ \gamma = 1-\delta & \text{if } \delta \le \triangle \\ \alpha = \dfrac{\delta}{(1+\delta)},\ \gamma = \dfrac{1}{1+\delta} & otherwise \end{cases} \qquad (12)$$

$\triangle$은 주어진 요구 간섭률을 의미한다. (그림 6)은 $\triangle = 0.1$ 일 때 α와 γ의 변화를 보여준다. 그림에서 보는 바와 같이 $\triangle = 0.1$을 기준으로 α와 γ가 위치를 바꾼 상태로 대칭을 이루는 것을 볼 수 있다. 따라서 Q-learning에 상대적으로 영향력이 큰 α는 δ가 $\triangle$보다 커질 때 γ 보다 작은 값을 가짐으로써 $\triangle$ 값을 유지하는 범위에서 기회주의적으로 채널접근을 시도하게 해준다.

4.2.4 알고리즘 구현

Q-learning 기반 채널접근기법은 각 CR 기기상에서 동작하며 초기화 단계로서 α=γ=0.5 , 모든 $Q(s_i,a_j)$, $i=j=\{0,1\}$ 는 0과 1 사이의 무작위 값으로 초기화된다. 초기 상태 $s(0)$ 및 그에 상응하는 행동 $a(0)$를 설정하기 위하여 ED를 사용하여 스펙트럼 센싱을 수행한 후 그 결과를 초기 상태 $s(0)$에 할당하고 할당된 s_i가 H_0일 경우에는 a_0를, H_1일 경우에는 a_1을 할당한다. 다음으로 알고리즘은 사전 학습 단계로 넘어간다. 사전 학습 단계에서는 ED의 스펙트럼 센싱 결과에 기반하여 Q-learning을 수행하지만 Q-learning의 결과를 직접 사용하지 않고 오직 학습만 수행한다. 사전 학습 단계는 다음과 같다. 먼저 앞서 결정된 쌍 $(s(0),a(0))$에 대한 보상 값을 계산하기 위해 실제 상태(true state)를 판단한다. 실제 상태는 다른 CR 기기로부터 수신 받아야 할 SM, RTS/CTS, 데이터, ACK 패킷이 수신되지 않으면 해당 채널에 PU가 활동하는 것(=S1)으로 판단되고, 정상적으로 수신 받아야 할 패킷들을 수신 받게 되면 PU가 활동하지 않는 것(=S0)으로 판단된다. 이를 기반으로 (표 1)을 참조하여 쌍 $(s(0),a(0))$의 보상 값 $r(s,a)$를 계산한다. 다음으로 새로운 상태 $s(1)$을 얻기 위해 스펙트럼 센싱을 수행하고 만약 그 결과가 상태 s_i이고 Q-learning은 $(1-\epsilon)$ 확률로 exploitation 모드 일 경우 $a(1)=\max_{a_j \in A}(Q(s_i,a_j))$를 만족시키는 행동 a_j가 선택되고, ϵ 확률로 exploration 모드 일 경우에는 선택 가능한 행동이 a_0와 a_1뿐이므

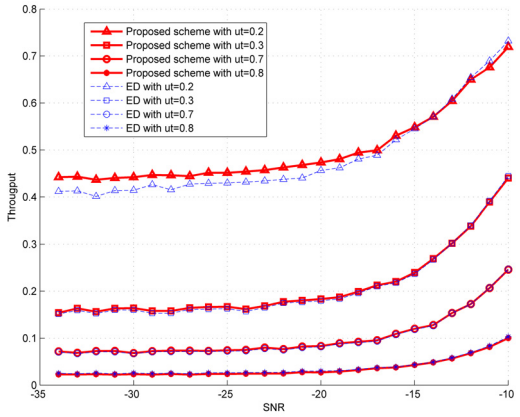


(그림 7) Q-learning 기반 채널접근기법의 전체 흐름도

로 $a(1)=\min_{a_j \in A}(Q(s_i,a_j))$를 만족시키는 행동 a_j 가 선택된다. 그리고 $Q(s(0),a(0))$ 값을 갱신하기 위하여 수식(12)과 수식(11)이 차례로 수행된다. 그리고 사전 학습의 첫 단계인 보상 함수를 이용한 보상 값 계산으로 다시 돌아간다. 이러한 사전 학습 단계는 Q-learning의 결과로 나온 행동 $a(t)$를 기반으로 채널접근을 시도했을 때 얻어질 것으로 예상되는 δ가 $\delta \le \triangle$ 을 만족할 때까지 수행되고, 조건을 만족 할 시 실제로 Q-learning의 결과를 채널접근에 적용하는 Q-learning 기반 채널접근 단계로 넘어간다. Q-learning 기반 채널접근 단계는 사전 학습 단계와 동일한 순서로 동작한다. 다만 실제로 Q-learning의 결과를 실행한다는 점이 다르다. 따라서 이상의 알고리즘은 (그림 7)의 순서도로 표현될 수 있다.

5. 모의실험

분산형 CR 네트워크에서 제안하는 Q-learning 기반의 채널접근기법의 성능을 확인하기 위하여



(그림 8) AWGN 채널에서의 처리율 성능 비교

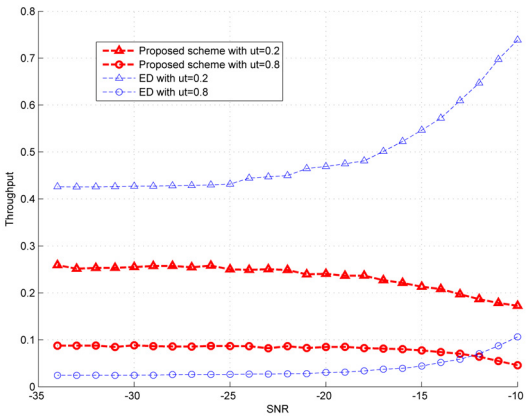
ON-OFF 모델로 동작하는 PU와 정의된 MAC 프로토콜에 의해 이미 동기화된 미니 슈퍼프레임 상에서 동작하는 CR 기기들을 고려한다. CR 네트워크상에 사용되는 채널은 하나로 고정하고, AWGN 채널 모델과 다음과 같이 지수분포를 가지는 레일레이(Rayleigh) 페이딩 채널 모델을 사용한다.

$$f(x)=\frac{1}{\bar{x}}exp\left(-\frac{x}{\bar{x}}\right), \quad x \ge 0 \qquad (13)$$

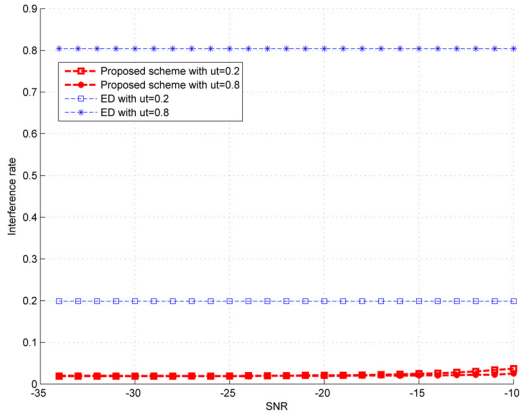
x는 SNR을, $\bar{x}$는 평균 SNR을 의미한다.

ED의 스펙트럼 센싱은 대역폭 4 Mhz에서 50 us의 길이로 이루어진다(M=200). Q-learning의 초기의 α와 γ는 0.5를 할당하고, 주어진 간섭률은 $\triangle$=0.1 이라 가정한다. 성능 평가 기준은 주어진 $\triangle$ 값을 만족하는 범위 안에서 PU 신호의 SNR 값에 따른 처리율 이며, Proposed scheme로 표기된 제안된 기법을 적용한 에너지 검출기와 ED로 표기된 일반적인 에너지 검출기를 비교 분석한다.

(그림 8)은 AWGN 채널 하에서 주어진 간섭률을 만족하는 경우의 처리율 성능을 SNR 값에 따라 비교한 모의실험 결과이다. ut 는 모의실험 중에 얼마나 PU가 채널을 점유했는지를 나타내는 채널 이용률(utilization)을 의미한다. ED는 AWGN



(그림 9) AWGN+레일레이 채널에서 처리율 성능 비교



(그림 10) AWGN+레일레이 채널에서 간섭률 성능 비교

채널 하에서 주어진 간섭률을 만족하는 수식(4)의 임계값 λ에 기반하여 동작하므로 Q-learning은 ED에 의해 결정된 채널의 현재 상태를 따르는 것이 상대적으로 높은 보상값을 얻을 것임을 학습 과정을 통하여 판단한다. 따라서 제안된 기법은 ED와 거의 유사한 형태의 처리율 결과를 보여준다. 하지만 ut=0.2인 경우, -15dB 이후로는 더 나은 성능을 보여주는데 이것은 Q-learning이 PU의 낮은 채널 이용률로 인해 발생하는 간섭 위험 수준의 저하를 수식(12)에 기반하여 더 많은 기회주의적 채널접근기회로 바꿀 수 있기 때문이다.

(그림 9)와 (그림 10)은 AWGN과 레일레이 페이딩 채널 하에서 처리율과 간섭률 성능을 SNR 값의 변화에 따라 비교한 모의실험 결과들이다. 그림 9에서 ut=0.2 인 경우, ED는 제안된 방식에 비해 높은 처리율 성능을 가지지만 그림 10에서 보는 바와 같이 간섭률이 상대적으로 매우 높아 PU의 통신을 거의 보장해 줄 수 없다. ut=0.8 의 경우, 제안된 방식은 -13dB를 기점으로 ED 보다 더 나은 처리율 성능을 보여주는 동시에 더 낮은 간섭률을 보장한다. 이는 PU의 잦은 채널 점유로 인해 PU의 채널 사용 패턴을 더욱 쉽게 Q-learning이 학습 할 수 있기 때문이며, 동시에 SNR 값이 낮아질수록 ED에 의해 결정되는 현재 상태가 점점 신뢰도를 상실함으로써 상대적으로 학습결과에 더 민감하게 반응하여 채널에 접근하기 때문이다. 이러한 결과로부터 PU의 채널 사용률이 일정수준 이상으로 높아질 경우 Q-learning이 일정 기간의 학습 후 현재 상태에 거의 상관없이 채널에 기회주의적으로 접근함으로써 더 나은 처리율 및 더 낮은 간섭률을 제공해 줄 수 있음을 예상할 수 있다.

6. 결 론

본 논문에서는 분산형 CR 네트워크에서 ED를 사용하여 PU의 채널사용을 검출하는 CR 기기의 스펙트럼 센싱 성능을 향상시키기 위하여 Q-learning 기반의 채널접근기법을 제안하였다. 제안하는 기법은 주어진 PU에 대한 간섭률을 보장하는 범위에서 처리율 성능을 향상시키기 위하여 Q-learning의 학습 결과를 ED의 스펙트럼 센싱에 적용하였다. CR 기기의 시스템 초기화 이후 사전 학습(Pre-play learning) 단계에서는 ED의 센싱 결과를 그대로 사용하고, 사전 학습 단계에서 지정된 간섭률의 임계값 이하로 학습이 진행되면 Q-learning 기반 채널접근(Q-learning normal operation) 단계를 넘어가 실제 학습의 결과를 기반으로 PU의 채널 사용유무를 판단하도록 하였다. 모의실험을 통하여 AWGN 및 레일레이 페이

딩 채널 하에서의 처리율 및 PU에 대한 간섭률 성능을 일반적인 ED와 비교 분석함으로써 성능 이득을 검증하였다. 특히, 레일레이 페이딩 채널 하에서 PU의 채널 사용률이 높을수록 처리율 및 PU에 대한 간섭률 성능이 ED보다 훨씬 더 높은 성능이득을 가짐을 확인 할 수 있었다.

감사의 글

This work was supported by KRF, funded by MEST (Mid-Carrier Researcher Program 2010 - 0009661)

참 고 문 헌

[1] 김창주, 임차식, “Cognitive Radio 기술 및 표준화 동향”, 한국전자파학회, 전자파기술, 제19권, 제2호, pp. 23-29, 2008년 3월

[2] 김재명, “Cognitive Radio 기술개요 및 발전방향”, 대한전자공학회, 전자공학회지, 제36권 제6호, 20-27쪽, 2009년 6월

[3] 고광진, 황성현, 정병장, 김창주, “Cognitive Radio 기술 및 표준화 동향”, 한국통신학회, 한국통신학회지, 제27권, 제8호, pp. 1-7, 2010년 7월

[4] Ian F. Akyildiz, Won-Yeol Lee, Kaushik R. Chowdhury, “CRAHNs: Cognitive radio ad hoc networks”, Ad Hoc Networks, Volume 7, Issue 5, pp. 810-836, July 2009

[5] Nguyen-Thanh Nhan, Xuan Thuc Kieu, Insoo Koo, “Cooperative Spectrum Sensing Using Enhanced Dempster-Shafer Theory of Evidence in Cognitive Radio”, Lecture Notes in Computer Science, Volume 5755, pp. 688-697, 2009

[6] Zhi Quan, Shuguang Cui, Poor H., Sayed A., “Collaborative Wideband Sensing for Cognitive Radios”, IEEE Signal Processing Magazine, Volume 25, Issue 6, pp. 60-73, 2008

[7] Junhong Nie, Haykin, S., “A Q-learning-based dynamic channel assignment technique for mobile communication systems”, IEEE Transactions on Vehicular Technology, Volume 48, Issue 5, pp. 1676 - 1687, 1999

[8] Mo Li, Youyun Xu, Junquan Hu, “A Q-Learning based sensing task selection scheme for cognitive radio networks”, Wireless Communications & Signal Processing 2009, pp. 1-5, 2009

[9] Fangwen Fu, van der Schaar M., “Learning to Compete for Resources in Wireless Stochastic Games”, IEEE Transactions on Vehicular Technology, Volume 58, Issue 4, pp. 1904-1919, 2009

[10] Yau K.-L. A., Komisarczuk P., Teal P. D., “ A context-aware and Intelligent Dynamic Channel Selection scheme for cognitive radio networks”, Cognitive Radio Oriented Wireless Networks and Communications 2009, pp. 1-6, 2009

[11] Tom Mitchell, “machine learning”, McGraw Hill, pp. 397-383, 1997

◐ 저 자 소 개 ◑

이 영 두

2007년 울산대학교 전기전자정보시스템 공학부 학사 졸업.
2009년 울산대학교 전기전자정보시스템 공학부 석사 졸업.
2009년~현재 울산대학교 전기전자정보시스템 공학부 박사 과정.
관심분야 : 차세대 이동통신, 무선센서네트워크, 무선인지 시스템
E-mail : leeyd1004@naver.com

구 인 수

1996년 건국대학교 전자공학과 학사 졸업.
1998년 광주과학기술원 정보통신공학과 석사 졸업.
2002년 광주과학기술원 정보통신공학과 박사 졸업.
2002년~2004년 광주과학기술원 연구교수
2002년~2003년 스웨덴왕립공과대학, 박사후 연수과정
2005년~현재 울산대학교 교수
관심분야 : 차세대 이동통신, 무선센서네트워크, 무선인지 시스템
E-mail : iskoo@ulsan.ac.kr