POMDP와 Exploration Bonus를 이용한 지역적이고 적응적인 QoS 라우팅 기법

정회원 한 정 수*

A Localized Adaptive QoS Routing Scheme Using POMDP and Exploration Bonus Techniques

Jeong-Soo Han* ***Regular Member***

요 약

본 논문에서는 Localized Adaptive QoS 라우팅을 위해 POMDP(Partially Observable Markov Decision Processes)와 Exploration Bonus 기법을 사용하는 방법을 제안하였다. 또한, POMDP 문제를 해결하기 위해 Dynamic Programming을 사용하여 최적의 행동을 찾는 연산이 매우 복잡하고 어렵기 때문에 CEA(Certainty Equivalency Approximation) 기법을 통한 기댓값 사용으로 문제를 단순하였으며, Exploration Bonus 방식을 사용해 현재 경로보다 나은 경로를 탐색하고자 하였다. 이를 위해 다중 경로 탐색 알고리즘(SEMA)을 제안했다. 더욱이 탐색의 횟수와 간격을 정의하기 위해 ϕ와 κ 성능 파라미터들을 사용하여 이들을 통해 탐색의 횟수 변화를 통한 서비스 성공률과 성공 시 사용된 평균 홉 수에 대한 성능을 살펴보았다. 결과적으로 ϕ 값이 증가함에 따라 현재의 경로보다 더 나은 경로를 찾게 되며, κ 값이 증가할수록 탐색이 증가함을 볼 수 있다.

Key Words : Localized QoS Routing, Dynamic Programming, Reinforcement Learning, Markov Decision Process(MDP), Partially Observable Markov Decision Processes(POMDP), Exploration Bonus

ABSTRACT

In this paper, we propose a Localized Adaptive QoS Routing Scheme using POMDP and Exploration Bonus Techniques. Also, this paper shows that CEA technique using expectation values can be simply POMDP problem, because performing dynamic programming to solve a POMDP is highly computationally expensive. And we use Exploration Bonus to search detour path better than current path. For this, we proposed the algorithm(SEMA) to search multiple path. Expecially, we evaluate performances of service success rate and average hop count with ϕ and κ performance parameters, which is defined as exploration count and intervals. As result, we knew that the larger ϕ, the better detour path search. And increasing κ increased the amount of exploration.

Ⅰ. 서 론

현재 많은 연구에서 Localized QoS 라우팅 기법이 Global QoS 라우팅 기법과 비교하여 보다 안정적이고, 간단하며, 네트워크 상황에 보다 적응적이라는 것이 제시되고 있으며[1], 이에 대한 일환으로 [2][3]에서는 Proportional Sticky Routing(PSR) 기법이 소개 되었다. 이러한 PSR 기법은 허용 가능한

※본 연구는 한국학술진흥재단 신진교수연구과제지원사업(KRF-2003-003-D00200) 지원으로 수행되었습니다.
* 신구대학 인터넷정보과 (jshan@shingu.ac.kr)
논문번호 : KICS2006-01-033, 접수일자 : 2006년 1월 18일, 최종논문접수일자 : 2006년 2월 17일

최대 데이터 블러킹 파라미터를 사용하여 각 사이클 당 경로를 따라 전송하게 될 플로우의 양을 제어함과 동시에 각 경로에 할당하게 될 데이터 비율을 재조정함으로써 자체 적응력을 유지하는 기법을 제공하고 있다. 그러나 여기에는 세 가지 문제점을 가지고 있다. 첫째, 실제 네트워크에서 트래픽 패턴을 항상 알 수 있는 것은 아니다. 두 번째로, 블러킹 확률을 계산하기 위해 경로 상의 정확한 정보를 알아야 한다는 것이다. 마지막으로, 비록 데이터의 패턴과 블러킹 확률에 대한 계산이 가능할 지라도 경로 전체에 대한 최적화 문제(global optimization problem)를 해결하기 위해 소요되는 시간은 상당히 크다고 할 수 있다[4].

이에 [4]에서는 강화학습(Reinforcement Learning)과 같은 지능적인 제어 방식을 이용하여 전체 네트워크에 대한 정보나 네트워크의 트래픽 패턴을 알지 못해도 지역적 라우팅이 가능한 Q-Learning 기반의 경로 선택 기법을 제안했다. 이러한 강화학습 기법은 에이전트(agent)와 환경(environment)과의 상호관계와 이에 따른 강화신호(reinforcement signal)를 통하여 에이전트의 행동을 개선해 나가는 방법으로서 환경에 대한 정확한 사전 지식 없이 학습 및 적응성을 보장할 수 있는 방법이다. 이러한 강화학습에서의 환경은 일반적으로 finite-state Markov Decision Process(MDP)로서 형식화되고, 이러한 맥락을 위한 강화학습 알고리즘은 dynamic programming 기술과 직접적으로 관련된다. 그러나 에이전트와 연결된 환경은 계속해서 변화하게 되어 현재 최적의 행동(action)이 미래에 그대로 보장되지는 않는다. 이에 환경에 대한 에이전트의 불확실성(Uncertainty)의 확률적 접근이 필요하게 되는데, 이러한 연구가 POMDP이다. 이 POMDP에서 사용하는 CEA 기법은 환경이 nonstationary 한 상태에서 최적의 결론(optimal solution)을 얻어내기 어렵기 때문에 사용하는 방법이다. 현재까지 MDP 환경 하에서 학습을 통한 라우팅 기법 연구는 소개되고 있으나 POMDP를 적용한 라우팅 기법에 대한 연구는 없는 실정이다.

따라서 본 논문에서는 네트워크의 상태가 계속해서 변화되는 환경 하에서 POMDP를 적용한 CEA 기법을 통해 지역적이고 적응적인 라우팅 기법을 소개하고자 한다. 이 기법을 적용하기 위해 네트워크 상에서 다중 경로를 탐색하는 알고리즘을 제시하였다. 이를 통해 경로 선택 시 사용되는 탐색 기법으로 Exploration Bonus 기법을 적용하여 오래 동안 사용되지 않은 경로에 더 많은 탐색 기회를 제공함으로써 전체적인 네트워크 성능을 향상시키고자 하였으며, 이를 위한 Dynamic Programming을 새롭게 정의하고 이를 통해 다양한 성능 파라미터를 도출 하였다.

Ⅲ. POMDP 기반의 지역적이고 적응적인 라우팅 기법

지역적 라우팅 기법은 네트워크 상태를 정확히 알지 못하는 상태에서 단지 송신지에서 유지하는 정보만을 의존하여 라우팅하게 되며, 이러한 상황은 결국 POMDP하에서 에이전트가 환경에 대한 관찰을 통해 결정하게 되는 방식과 연결될 수 있다. 즉, 에이전트는 첫째, 매 시도(t)시에 자신의 상태($x^t(n) \in X$)에 대한 정보와 두 번째, 각 행동을 취할 시 비용에 대한 정보($C_x(a)$), 세 번째 최종 목적지 상태에 대한 정보, 마지막으로 전이 확률에 대한 확률분포에 대한 정보들은 알 수 있지만 전체 네트워크 상태 정보를 알 수 없기 때문에 특정 행동에 대한 성공 확률은 알 수 없다. 이러한 문제로 인해 지역적 라우팅 문제인 MDP을 POMDP로 적용할 수 있다. 또한 네트워크 상황에 적응적 라우팅 기법은 환경에 대한 관찰과 이에 대한 갱신을 통한 에이전트 대응 방식과 연결될 수 있다.

3.1 시스템 모델

3.1.1 POMDP 기반의 라우팅 모델

본 논문에서 제안한 라우팅 모델을 사용하기 위해 [표 1]에서와 같이 네트워크 라우팅 항목과 POMDP상의 항목에 대한 상호 연결이 필요하다.

또한, POMDP 모델을 적용하여 송신지에서 전체 네트워크 상태 정보에 대한 불확실성으로 인해 목적지로 가는 다중 경로 중에 각각의 경로 선택에 대한 서비스 성공률을 확률변수로 표현하기로 한다. 즉, 특정 상태 x에서 행동 a를 취하여 상태 $y \in X$로 성공적으로 전이할 확률을 $p_{xy}^t(a)$(단, t는 시도

표 1. 라우팅 항목과 POMDP 항목 연결.
Table 1. The mapping of routing element and POMDP

네트워크 라우팅 항목	POMDP 기법 항목
각 네트워크 노드	에이전트
목적지로의 경로들 집합	행동들($a \in A$)
요청을 수신한 노드	상태($(x \in X)$
지역 라우팅 정책	최적화 정책($V_{\alpha+1}^{t,n}(x)$)

횟수)로 정의하기로 한다. POMDP 모델에서 사용하는 전이확률은 시간이 지남에 따라 그 값이 변하게 되는 non-stationary 네트워크 환경 특성으로 인해 이들의 확률분포 $U[P_{xy}^{t+1}(a)|P_{xy}^{t}(a)]$를 기반으로 Markov 방식에 따라 변하게 된다. 따라서 특정 행동 a에 대한 확률분포(신뢰상태) 값은 이미 경험한 확률값(이전 값)과 이들의 확률분포를 기반으로 하여 Baye's rule를 통해 결정되며, 갱신하게 되며 이렇게 결정된 값들은 라우팅 정책에 이용된다. 결국 에이전트는 실제적인 전이확률 $p_{xy}^{t}(a)$은 알지 못하지만, 환경을 통해 얻어진 데이터를 통해서 전이확률들에 대한 확률분포만 알게 되고 이를 통해 전이확률을 유추할 수 있다. 또한, 본 논문에서는 POMDP 모델과 함께 에이전트의 적당한 행동을 결정하기 위해 확률변수를 사용하는 대신에 이들의 평균값 $q_{xy}^{t,n}(a) = E_{t^{t,n}}[p_{xy}(a)]$을 사용하는 CEA 기법을 사용하고자 하는데, 이는 POMDP 문제를 해결하기 위해 Dynamic Programming을 사용하여 최적의 행동을 찾는 연산이 매우 복잡하고 어렵기 때문에 CEA 기법을 통한 기댓값 사용으로 문제를 단순화하고자 함이다. 여기서 $\tau^{t,n}[p_{xy}(a)]$는 t시도가 발생된 n시간의 전이확률에 대한 확률분포를 나타낸다. 그리고 특정 상태 x에서 행동 a를 취할 때의 비용(다시 말해, 송신지에서 목적지까지의 홉 수)를 $C_x(a)$로 정의한다. [그림 1]은 본 논문에서 사용한 POMDP와 CEA 기법을 네트워크 환경에 적용한 것이다. 이를 토대로 에이전트의 정책은 합산된 비용의 기댓값이 최소가 되는 행동 a를 결정하는 것으로, 근사값을 이용하여 (6)과 같은 반복 연산을 수행하여 결정하게 된다.

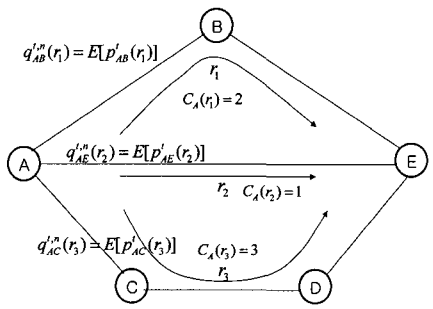

그림 1. POMDP와 CEA 기법을 적용한 네트워크 환경 모델
Fig. 1. The Network environment model with POMDP and CEA techniques

$$V_{\alpha+1}^{t,n}(x) = min_{a \in A}\{C_x(a) + \gamma \sum_{y \in Y} q_{xy}^{t,n}(a)\, V_{\alpha}^{t,n}(y)\} \tag{6}$$

여기서, $0 \le \gamma \le 1$은 할인율을, α는 반복횟수를 나타낸다.

3.1.2 라우팅 정보 갱신

라우팅 정보 갱신은 에이전트가 식(6)을 통해 여러 개의 경로 중 기댓값이 최소가 되는 행동 a(경로)를 선택하게 되면 그 후에 수행하게 된다. 즉 (6)의 오른쪽 수식이 최소가 되는 행동을 선택하고 이를 통해 라우팅하게 된다. 그리고 나서 실제 관찰을 통해 전이확률에 대한 새로운 확률분포 $q_{xy}^{t,n+1}(a)$를 계산하게 된다. 목적지까지의 여러 경로 중 각 경로상의 QoS(잔여대역폭 등)값이 사용자 서비스가 요구하는 QoS 값보다 크거나 같으면 해당 경로에 대한 라우팅이 성공하게 되는데 본 논문에서는 그 경로에 대해 라우팅이 유효(efficacy)하다라고 표현하고 $e_x^{t,n}(a) = 1$로 정의한다. 반대로 QoS 값을 만족하지 못하여 라우팅이 실패한 경로에 대해서는 유효하지 못하다(inefficacy)라고 표현하고 $e_x^{t,n}(a) = 0$으로 정의한다. 본 논문에서는 전체 네트워크에 대한 정보를 알지 못한 상태에서 다중 경로에 대한 라우팅 성공률을 알 수 없기 때문에 $e_x^{t,n}(a)$ 값을 정의할 수 없다. 따라서 라우팅 성공률 즉, 각 경로에 대한 전이확률을 확률적 모델(ϕ)로 정의하고자 한다. 이에 $q_x^{t,n}(a)$를 상태 x에서 t시도 동안 n시간의 행동 a에 대한 유효확률로써 $e_x^{t,n}(a) = 1$인 확률을 말한다. 이와 함께 각 사용자 서비스들의 트래픽 패턴을 알 수 없기 때문에 이들의 평균 도착율(κ)을 함께 적용하여 새로운 전이 확률분포를 갱신하게 된다.

즉, $t-1$ 시도 후에 에이전트의 다음 확률분포 값($q_x^{t,0}(a)$)은 (7)과 같이 갱신함으로써 라우팅 정보를 갱신하고자 한다.

$$q_x^{t,0}(a) = \begin{cases} \kappa\phi + (1-\kappa) q_x^{t-1,0}(a) & (7-1) \\ 1 - k(1-\phi) & (7-2) \\ k\phi & (7-3) \end{cases} \tag{7}$$

여기서, (7-1)은 사용자 서비스의 라우팅 시도 시 상태 x에서 행동 a를 시도하지 않았을 때이며, (7-2)는 상태 x에서 행동 a가 시도되어 성공했을 때의

갱신 값을 나타낸다. 또한, (7-3)은 상태 x에서 행동 a가 시도되어 성공하지 못했을 때의 갱신 값을 나타낸다.

(7)에서 보는 바와 같이 $q_x^{t,n}(a)$은 에이전트 상태 x에서 행동 a에 대해서 어떠한 결과가 발생되더라도 다른 값으로 설정이 될 것이다. 즉, 시도되지 않은 행동에 대해서는 κ의 갱신 확률로 ϕ값으로 접근할 것이고, 성공하거나 실패했을 경우는 각각 해당 값으로 재설정이 될 것이다. 단, 라우팅 초기에 각 에이전트들은 각 경로들에 대한 전이확률이 유효할 확률 ϕ를 가지고 있다고 가정한다($q_x^{1,0}(a)=\phi,\ \forall x,a$).

3.1.3 경로선택 기법(Exploration Bonus 기법)

지역적 라우팅 환경 하에서의 네트워크 상의 각 노드(에이전트)들은 네트워크 전체 정보에 대한 지식이 없기 때문에 목적지로의 각 경로들 상의 서비스 성공 확률(전이확률)을 알지 못한다. 이에 각 노드들은 라우팅 초기에는 임의의 초기값을 가지고 서비스하게 되며, 네트워크 환경이 계속적으로 변화된다는 가정 하에 각 노드들은 탐색(exploration)을 하게 된다. 그러나 각 노드들이 자신이 네트워크 전체 정보와 각 경로들 상의 서비스 성공률을 알지 못한다는 것을 알지라도 다중 경로 상에 정확한 시도를 한다는 것은 매우 어려운 일이다. 이에 본 논문에서 제시하는 Exploration Bonus 기법은 계속적인 환경 변화 상태에서 문제를 해결하기 위한 heuristic 방법으로 제시될 수 있다.

Sutton[5]은 움직이는 장벽이 있는 미로 상에서 Exploration Bonus 기법을 통해 목적지로 도달하는 실험을 통해서 선택된 횟수가 적거나 오래 전에 선택된 행동에 대해 보너스를 적용함으로써 그 성능을 높이는 결과를 보여주고 있다.

본 논문에서는 각 경로들에 대한 전이확률에 대해 평균값을 사용하는 CEA 기법을 사용한다. 그러나 이러한 CEA 기법은 목적지로 가는 하나의 경로 상에 QoS 문제로 인해 서비스가 블러킹 된다면 에이전트가 그 목적지로의 다른 경로를 선택하지 못할 것이다. 따라서 이러한 문제를 해결하기 위한 heuristic 방법으로 Exploration Bonus 기법을 사용하고자 한다.

이미 선택된 횟수가 적거나 오래 전에 선택된 행동에 보너스를 적용하여 라우팅 성능을 평가한 연구가 있다[6]. 이에 본 논문에서는 오래 전에 선택된 행동에 보너스를 적용한 시간의 함수를 사용하는 연구를 시행한다.

각 경로들에 대한 전이확률의 효율성 모델을 사용함과 동시에 만약 그 행동이 $n_x(a)$ 시도동안 시도되지 않았을 경우 그 행동에 Exploration Bonus 값인 $\alpha\sqrt{n_x(a)}$를 직접적으로 추가하는 보너스 기법을 사용한다($\alpha>0$). 여기서 α 값은 탐색의 횟수를 제어하는 데 사용될 수 있다.

3.1.4 성능 파라미터(ϕ, κ)

상태 x에서 목적지로 가는 행동 a를 취했을 경우 상태 y로 가게 되고 이러한 행동이 유효하다고 가정하자. 또한, 목적지로 가기 위해 상태 x 상에 존재하는 다른 경로들을 기준으로 상태 y가 목적지에 더 가깝다고 가정해 보자. 즉, $V_\alpha^{t,n}(x) \gg V_\alpha^{t,n}(y)$일 때 $q_x^{t,n}(a)$값이 커질수록, 수식 (6) 상에서 행동 a에 대한 오른쪽 수식 상의 비용 값($C_x(a)$)은 더 작아질 것이다. 이것은 $V_{\alpha+1}^{t,n}(x)$의 비용 값이 더 작아지는 가능성을 가지게 된다. 그러므로 각 노드에서는 목적지로 가기 위해 환경에 적응적인 다른 행동들을 계속적으로 탐색하게 될 것이다.

또한, 수식 (7)에서 보듯이 해당 행동이 시도된 후 시간이 증가함에 따라 $q_x^{t,n}(a)$값이 ϕ 값으로 접근하게 된다. 이는 다시 말해서 ϕ 값이 커질수록 위에서 제시한 Exploration Bonus를 적용한 것과 같게 된다는 것이다.

이와 같이 ϕ 파라미터는 각 노드에서 exploration과 exploitation 사이의 균형을 제어하는데 사용된다. 즉, ϕ=1이면 각 노드는 현재 경로보다 더 짧은 경로를 찾기 위해 탐색을 하게 된다. 이것은 각 노드에서 최단 거리 찾기 문제와 일치하게 된다. 반대로 ϕ값이 작아질 수록 각 노드는 현재의 경로보다 더 좋은 detour를 찾지 않을 것이다. 또 다른 κ 파라미터는 exploration과 exploitation 사이의 양에 대해 간접적으로 영향을 제공한다.

다시 말해 ϕ 파라미터 값은 현재보다 더 나은 경로를 찾기 위한 탐색의 양에 영향을 미치고, κ 파라미터 값은 탐색이 얼마나 자주 발생하는지를 제어하는데 영향을 미치게 된다.

Ⅱ. 다중 경로 탐색 기법

POMDP 상의 근사값을 이용한 라우팅 모델에서 사용할 다중 경로에 대해 본 논문에서는 S.Banerjee[9]

가 제안한 SSP(Single-Sink Program) 알고리즘을 변경한 SEMA(Shortest Edge-disjoint Multi-path searching Algorithm) 알고리즘을 제안한다. SEMA 알고리즘은 [12]에서와 같이 Dijkstra의 최단 거리 알고리즘을 반복적으로 사용하여 송신지에서 목적지까지의 최소 가중치를 갖는 여러 개의 edge-disjoint 경로들을 찾는 것이다.

그래프 $G=(V,E)$는 각 회선 $(u,v)\in E$에 비용 $c(u,v)$를 갖는 방향성 그래프로 하고 $|V|=n$, $|E|=m$로 정의할 때 SEMA 알고리즘은 [표 2]와 같다.

SEMA 알고리즘은 edge-disjoint한 다중 경로를 찾기 위해 $\Theta(n,m)$의 Dijkstra 알고리즘을 여러 번 실행하게 되므로 순차적인 $\Theta(n,m)$ 시간으로 해결할 수 있다.

표 2. 다중 경로 탐색 기법(SEMA) 알고리즘
Table 2. Shortest Edge-disjoint Multi-path searching Algorithm

SEMA(Shortest Edge-disjoint Multi-path searching Algorithm)

[초기화] 찾고자 하는 다중 경로 집합 $\delta=\{\phi\}$

[단계 1] 최단거리 계산(I)
송신지 s에서 Dijkstra 알고리즘을 통해 최단 거리 트리 T를 생성하고, T상의 s에서 목적지 v로의 최단 경로를 P_1으로 정한다. 또한, s에서 각 노드 x의 비용을 $C(s,x)$로 정의한다.

[단계 2] 비용 재계산 및 그래프 수정
G 상의 모든 회선(a,b)의 비용은 $c'(a,b)=c(a,b)+C(s,a)-C(s,b)$로 재계산된다. 이때, T에 속하는 모든 회선은 0으로 계산된다. 또한, P_1에 속하는 G 상의 회선들의 방향을 반대로 구성하여 새로운 그래프 G_v를 생성한다.

[단계 3] 최단거리 계산(II)
G_v상의 s에서 v로의 최단거리를 계산하고 이를 P_2로 정한다.

[단계 4] 최단경로들 생성
$\{P_1\cup P_2\}-\{P_1\cap P_2\}$ 결과인 P_1'과 P_2'를 생성하고 (이들이 disjoint shortest path이다), 다중 경로 집합 $\delta=\{P_1',P_2'\}+\delta$를 계산한다.

[단계 5] 그래프 축소
G 상에서 δ에 포함하는 회선들을 제외한 새로운 그래프 G'를 생성하고 이를 $G=G'$으로 재설정한다.

[단계 6] 반복
송신지 s에 연결된 회선이 없을 때까지 [단계 1]을 반복 수행한다.

Ⅳ. 시뮬레이션 및 결과 분석

본 절에서는 논문에서 제안하는 새로운 알고리즘들(POMDP, CEA)을 통해 Exploration Bonus를 적용한 결과와 함께 ϕ, κ 파라미터들에 의한 탐색 횟수에 대한 결과를 살펴보도록 하겠다. 이를 위해 다음과 같은 테스트 환경을 제공하기로 한다.

4.1 시뮬레이션 환경

[그림 2]는 본 논문에서 제안한 알고리즘들의 성능 평가를 위해 사용된 네트워크 토폴로지들을 보여주고 있다. 여기서 사용하는 성능 평가 환경은 [1][2][3]에서 사용된 시뮬레이션 환경을 그대로 사용하고 있다. 따라서 다음과 같은 가정을 사용한다. 먼저 모든 회선은 무방향성이고, 각 방향으로 똑같은 C unit의 대역폭을 갖는다. 네트워크에 도착한 연결 요청은 1 unit 대역폭을 요구한다고 가정하자. 연결 요청(평균 도착율)은 소스 노드에 κ를 갖는 포아송 프로세스를 따르며, 목적지는 소스 노드를 제외한 모든 노드로부터 랜덤하게 선택된다. 연결 요청에 대한 지속시간은 $1/\mu$를 갖는 지수 분포를 따른다. 네트워크 부하는 $\rho=\lambda Nh/\mu LC$를 정의한다. 여기서 N은 소스 노드의 총수이며, L은 회선의 총수, h는 평균적으로 모드 소스-목적지 쌍에서 연결 요청 당 평균 홉 수를 나타낸다. 또한 시뮬레이션에서 사용된 파라미터들을 C=20, μ=60 sec으로 정의한다. 소스 노드 상의 평균 도착율 κ는 성능 파라미터로 사용한다.

또한 연결 요청을 전송할 사용 가능한 경로(feasible path)는 위에서 살펴본 다중 경로 찾기 알고리즘(SEMA)을 통해 주 경로(primary path)와 보조 경로(alternative path)들로 분류하여 사용한다. 마지막으로 Exploration Bonus기법은 초기에 greedy-policy 기법을 사용하다가 연결 요청에 대한 블러킹이 발생했을 때 적용하게 된다.

4.2 결과 분석

사용자 서비스가 요청되면 에이전트 상에서 적절한 행동이 목적지에 도달 때까지 선택된다. 목적지에 도달하기 위한 행동들을 선택하기 위해서는 수식 (6)이 행동을 결정하기 위한 각 시도 전에 수행되어야 하며, 이를 통해 테스트 네트워크 환경 상에서 적절한 행동이 선택된다. 이러한 모델은 에이전트가 각 시도에 대한 선택된 행동이 유효한지 그렇지 않은지를 판단함으로써 갱신된다. 마지막으로 목

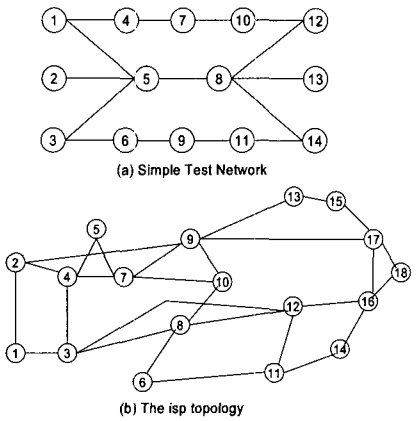

그림 2. 시뮬레이션 네트워크
Fig. 2. The simulation networks

적지에 도달하게 되면 기댓값 $q_x^{t,n}(a)$ 값이 갱신되게 된다.

[그림 3]에서는 ϕ값에 따른 테스트 네트워크별 라우팅 성공 시 평균 홉 수에 대한 결과를 보여주고 있다. 수식 (6)에서 알 수 있듯이 ϕ 값이 증가함에 따라 현재의 경로보다 더 나은 경로를 찾게 된다. 이는 현재의 경로가 다양한 QoS 파라미터들로 인해 서비스가 블러킹 되었을 경우 다른 경로들을 찾을 수 있도록 도와준다. 그러나 ϕ 값이 작아질수록 현재의 경로를 계속 사용하게 된다. 따라서 [그림 3]과 같이 결과가 나타난다. 즉, 현재의 경로보다 detour 경로를 찾게 되면 현재의 경로보다 홉 수가 증가하게 됨을 알 수 있다. 이는 ϕ 값이 증가함에 따라 현재의 경로보다 다른 경로를 탐색하기 때문이다. 이 때 사용되는 κ 값은 0.003으로 정한다.

[그림 4]에서는 ϕ값에 따른 테스트 네트워크별 서비스 성공률에 대한 결과를 보여주고 있다. 이 결과는 [그림 3]의 결과와 연관 지어 생각할 수 있다.

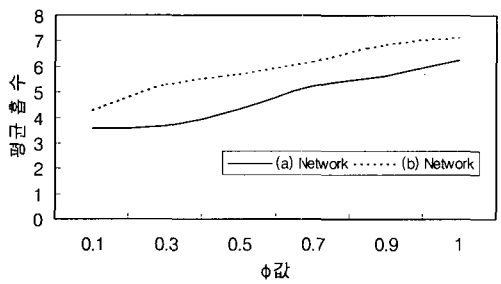

그림 3. ϕ값에 따른 테스트 네트워크별 라우팅 평균 홉 수
Fig. 3. The performance of routing average hop count per test network according to ϕ value

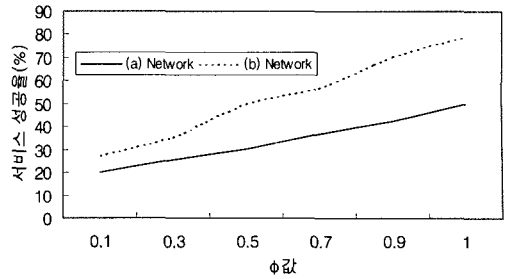

그림 4. ϕ값에 따른 테스트 네트워크별 서비스 성공률
Fig. 4. The performance of routing success rate per test network according to ϕ value

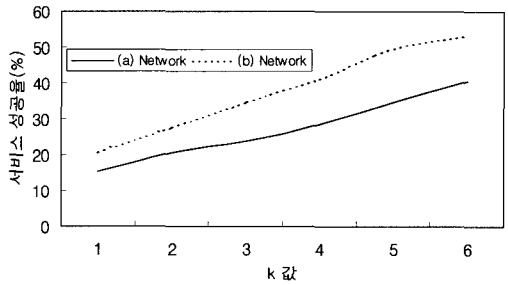

그림 5. κ값에 따른 테스트 네트워크별 서비스 성공률
Fig. 5. The performance of routing success rate per test network according to κ value

즉, ϕ 값이 증가 할수록 현재의 경로보다 detour 경로를 탐색하게 됨에 따라 서비스 성공률이 증가하게 될 것이다. 이 때 사용되는 κ 값 역시 0.003으로 정한다.

[그림 5]는 κ 값에 따른 테스트 네트워크별 서비스 성공률을 보여주고 있다. 앞서 살펴보았듯이 κ 값은 탐색이 얼마나 자주 발생하는지를 알 수 있는 파라미터이다. 이는 탐색이 자주 발생할 수록 현재의 경로보다 detour 경로를 찾게 되며, 이로 인해 서비스 성공률이 높아짐을 알 수 있다. 즉, κ 값이 증가할 수록 서비스 성공률이 높아진다. 이 때 사용된 ϕ 값은 0.5이다.

본 논문에서 제안한 알고리즘과 다른 연구에서 제시한 알고리즘에 대한 비교 분석을 보면 [그림 6]과 같다. 여기서 비교한 다른 연구는 [3]에서 제시한 psr Localized QoS Routing 기법과 [9]에서 제시한 TD Localized QoS Routing 기법이다. [그림 6]에서 사용된 psr 기법은 네트워크의 상황(트래픽 정보)를 알고 있어야 각 경로들의 blocking 확률을 계산할 수 있지만 실험을 위해 트래픽 정보를 알지 못하는 상태에서 임의로 계산하기로 한다. TD 기법은 [9]에

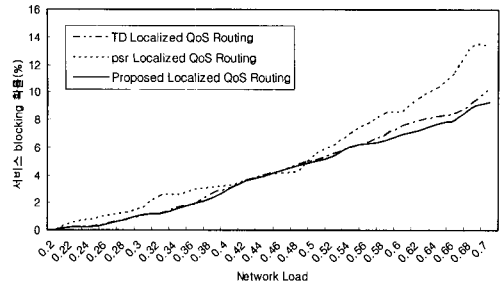

그림 6. (b)Network상에서 네트워크 부하에 따른 서비스 blocking 확률
Fig. 6. The performance of service blocking probability under various network loads on (b) network

서 사용한 값을 사용하고 본 논문에서 제시한 알고리즘을 위해 ϕ=0.5와 κ=0.003을 사용한다. [그림 6]에서 보는 바와 같이 TD 기법과 논문에서 제시한 알고리즘과 비슷한 성능을 보여주고 있는데 이는 네트워크 환경에 대한 확률적 접근과 Exploration Bonus를 사용하고 있기 때문이며, psr 기법은 환경에 대한 무지로 인해 성능이 떨어지고 있음을 알 수 있다.

Ⅵ. 결 론

본 논문은 네트워크 상의 전체 상태정보에 대한 사전 지식 없이 지역적으로 라우팅할 수 있는 새로운 라우팅 기법을 제안하였다. 제안된 라우팅 기법은 POMDP 기법과 이를 좀더 간단하게 적용하기 위해 CEA 기법을 제공함으로써 라우팅 기법을 지역적 라우팅 방식과 함께 네트워크 상황에 그 때 그 때 적응할 수 있는 적응적 라우팅 기법을 함께 제안하였다. 또한, 다중경로 탐색을 위해 SEMA 알고리즘을 제안했다.

POMDP 기법은 각 노드에서 네트워크에 대한 전체 상태정보를 알 수 없다는 지역적 라우팅 환경하에서 적용하기 위한 기법이며, CEA 기법은 POMDP 환경에서 적절한 행동을 선택하기 위해서는 Dynamic Programming을 통해 선택하게 되는데 이 때 사용되는 계산 비용과 오버헤드를 줄이기 위해 사용하였다. 또한 네트워크 상황에 맞게 유효성 값을 사용하기 때문에 네트워크 상태 정보를 간접적으로 사용하여 라우팅 할 수 있게 된다.

이렇게 네트워크 환경 상에서 지역적이고 적응적인 라우팅 기법을 제안하게 되었다. 하지만 경로 선택 문제에 있어서 네트워크 상황에 의해 목적지를 위한 경로가 항상 좋은 경로가 될 수 없기 때문에 이를 해결하기 위해 Exploration Bonus 기법을 적용하여 현재 경로 보다 더 나은 detour 경로를 탐색하고자 알고리즘을 추가하였다. 더욱이 탐색의 횟수와 간격을 정의하기 위해 ϕ와 κ 성능 파라미터들을 사용하여 이들을 통해 탐색의 횟수 변화를 통한 서비스 성공률과 성공시 사용된 평균 홉 수에 대한 성능을 살펴보았다.

결론적으로 ϕ 파라미터 값은 현재보다 더 나은 경로를 찾기 위한 탐색의 양에 영향을 미치고, κ 파라미터 값은 탐색이 얼마나 자주 발생하는지를 제어하는데 영향을 미치게 된다. 따라서 ϕ=1에 가까워지면 현재 경로보다 더 나은 detour 경로를 탐색하게 될 것이고, ϕ=0에 가까워 질수록 현재의 경로를 통해 서비스하고자 할 것이다. 따라서 ϕ=1에 가까워 질수록 서비스 성공률은 올라가게 되며, 평균 홉 수는 더 많아지게 됨을 알 수 있었다. 또한, κ 값이 올라갈 수록 탐색의 양이 증가하게 되므로 서비스 성공률이 높아짐을 알 수 있었다.

참 고 문 헌

[1] X.Yuan and A.Saifee, "Path Selection Methods for Localized Quality of Service Routing", *Technical Report*, TR-010801, Dept of Computer Science, Florida State University, July, 2001.

[2] Srihari Nelakuditi, Zhi-Li Zhang and Rose P.Tsang, "Adaptive Proportional Routing: A Localized QoS Routing Approach", *In IEEE Infocom*, April 2000.

[3] Srihari Nelakuditi, Zhi-Li Zhang, "A Localized Adaptive Proportioning Approach to QoS Routing", *IEEE Communications Magazine*, June 2002.

[4] Y.Liu, C.K. Tham and TCK. Hui, "MAPS: A Localized and Distributed Adaptive Path Selection in MPLS Networks" in *Proceedings of 2003 IEEE Workshop on High Performance Switching and Routing, Torino*, Italy, June 2003, pp. 24-28.

[5] Sutton, R.S. "Learning to predict by the method of temporal differences" *Machine Learning* 3. 1988, pp. 9-44.

[6] 한정수, "TD(λ) 기법을 사용한 지역적이며 적응적인 QoS 라우팅 기법" *한국통신학회* 제30권 제5B호 2005, pp 304~309

[7] Gregory Z. Grudic, Vijay Kumar, "Using Policy Gradient Reinforcement Learning on Automous Robot Controllers", *IROS03*, Las Vagas, US, October, 2003

[8] Richard S. Sutton etc, "Policy Gradient Methods for Reinforcement Learning with Function Approximation", *Advances in Neural Information Processing System,* pp. 1057~1063, MIT Press 2000

[9] S.Banerjee, R.K. Ghosh and A.P.K Reddy, "Parallel algorithm for shortest pairs of edge-disjoint paths",

한 정 수(Jeong-Soo Han) 정회원

1997년 2월 성균관대학교 정보공학과 졸업

1999년 2월 성균관대학교 전기전자및컴퓨터공학부 석사

2003년 2월 성균관대학교 전기전자및컴퓨터공학부 박사

~현재 신구대학 인터넷정보과 교수

<관심분야> 네트워크 관리, QoS라우팅, 서비스 복구 라우팅