사물인터넷융복합논문지 Vol. 12, No. 5, pp. 1-8, 2026 https://doi.org/10.20465/KIOTS.2026.12.5.001

대규모 IoT 접속 환경에서 셀 과부하 방지를 위한 AI 기반 지능형 기지국 파라미터 최적화 기법

박복녕[1*], 권준혁[2]
[1]LG유플러스 NW부문 팀장, [2]LG유플러스 NW부문 부사장

AI-based Intelligent Parameter Optimization Scheme for Preventing Cell Overload in Massive IoT Environments

Bok-Nyong Park[1*], June-Hyuk Kwon[2]
[1]Team Leader, NW Division, LG Uplus
[2]Executive Vice President, NW Division, LG Uplus

요 약 최근 IoT 기기의 급격한 증가로 인해 사물인터넷 전용망 (LTE Cat.M1, NB-IoT 등) 기반의 대규모 단말 접속 (Massive Connectivity)이 빈번해지고 있다. 특히 다수의 단말이 동시에 접속할 경우, 기지국 내 RRC(Radio Resource Control) 연결 자원이 고갈되어 신규 단말의 접속이 차단되는 과부하 현상이 발생한다. 이를 해결하기 위해 AI 기반 지능형 Inactivity Timer 최적화 알고리즘을 제안한다. 제안하는 최적화 알고리즘은 기존의 고정 임계치 방식이 가진 경직성을 극복하기 위해, 기지국 자원 가용성과 서비스 품질(절단율) 간의 관계를 수학적으로 정립하고 DQN(Deep Q-Network)을 통해 시스템 가중치를 실시간으로 최적화하였다. 성능 평가 결과, 제안된 AI 기반 지능형 하이브리드 제어 모델은 대규모 단말 접속 환경에서 기존 고정 방식 및 규칙 기반 대비 수용량을 향상시켰다.

주제어 : 사물인터넷 전용망, LTE Cat.M1, Inactivity Timer, AI, Hybrid Model

Abstract With the recent rapid proliferation of IoT devices, massive connectivity based on IoT network (LTE Cat.M1, NB-IoT and so on) is becoming increasingly frequent. In particular, when a large number of devices attempt to connect simultaneously, Radio Resource Control (RRC) connection resources within the base station become exhausted, leading to overload conditions that block new device access. To address this issue, this paper proposes an AI-based intelligent Inactivity Timer optimization algorithm. To overcome the rigidity of conventional fixed-threshold methods, the proposed algorithm mathematically defines the relationship between base station resource availability and Quality of Service (e.g., drop rate) and optimizes system weights in real-time using a Deep Q-Network (DQN). Performance evaluation results demonstrate that the proposed AI-based intelligent hybrid control model enhances capacity compared to existing fixed and rule-based methods in massive IoT environments.

Key Words : Cellular IoT,; LTE Cat.M1, Inactivity Timer, AI, Hybrid Model

*교신저자 : 박복녕(bnpark@lguplus.co.kr)
접수일 2026년 06월 01일 수정일 2026년 06월 28일 심사완료일 2026년 08월 19일

www.kci.go.kr

1. 서론

이동통신 환경은 LTE에서 5G, 그리고 6G로 진화하고 있다. 이와 같은 이동통신 기술이 발전하면서 다양한 서비스들에 통신 기술이 접목되며 IoT 서비스 기기들도 증가하고 있다[1, 2]. 특히 수도검침, AMI, 전기차 충전기, 공유 자전거 등과 같은 저용량의 대규모 서비스가 확대되면서 LPWA(Low Power Wide Area) 기술은 사물인터넷 전용망(LTE Cat.M1, NB-IoT 등)을 중심으로 massive IoT 환경으로 변하고 있다.

그러나, 5G가 상용화된 상황에서도 대용량 데이터 전송을 하는 M2M 기반의 broadband IoT만 상용화 되어 서비스되고 있으며, 수도검침, AMI, 전기차 충전기, 전기 자전거 등 저속의 통신 서비스에 제공되던 LPWA 기반의 서비스를 대체할 기술은 아직 상용화되지 않고 여전히 LTE 기반의 Cat.M1과 NB-IoT를 통해 서비스되고 있다. 그러다 보니, LPWA 서비스 속성의 massive IoT 단말은 LTE 환경에서 지속적으로 증가되고 있으며, 이로 인해 단일 셀 내 단말 밀도가 급격히 증가함에 따라 기지국의 RRC(Radio Resource Control) 연결 자원 관리는 네트워크 생존성과 직결되는 문제가 되었다[3]. 이러한 문제는 기지국 자원이 충분하다면 문제가 없지만, 한정된 기지국 자원으로 인해 통신을 원하는 모든 단말을 수용하기는 어려우며 이동통신사들의 LTE 투자 축소로 IoT 서비스를 위한 기지국 증설을 하기는 어려운 현실이다. 특히, Cat.M1의 경우 전체 기지국 RB(Resource Block) 중 6 RB 만을 할당 받아 일부 자원만 사용할 수 있기 때문에 이와 같은 문제는 더 해결하기 어려운 상황이다.

LTE Cat.M1과 같은 사물인터넷 전용망에서의 단말 초과접속에 따른 기지국 과부하 방지를 위해 기지국에서는 inactivity timer라는 특정 기지국 파라미터를 조절하여 데이터 전송이 끝난 단말의 접속해제 시간을 단축시키고 대기중인 단말 수를 줄여 고객서비스 품질을 향상 시킬 수 있다. 그러나, 현재 각 이동통신사에서는 고정된 값의 inactivity timer를 설정하여 사용함으로써 급격한 트래픽 변화에 유연하게 대응하지 못해 신규 접속 차단을 야기하는 등 한계를 드러내고 있다[4].

이와 같은 massive IoT 환경에서의 단말 수용 문제를 해결하고 변화하는 환경에 대응하기 위해 본 논문에서는 AI의 자율 학습 능력을 자원 관리에 적용함으로써, 수학적 선형화 모델과 인공지능의 자율 학습 능력을 결합하여 실시간으로 최적의 평형점을 도출하는 새로운 차원의 지능형 자원 관리 프레임워크를 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 이동통신의 사물인터넷 전용망(LTE Cat.M1) 환경에서 단말 초과 접속에 따른 문제와 조치 방안에 설명하고, 3장에서는 AI 기반의 지능형 inactivity timer 최적화 방안을 제안한다. 4장에서는 제안 방식의 AI 모델링과 실시간성을 분석하고, 5장에서는 제안하는 방식의 효과를 검증한다. 6장에서는 결론 및 추후 연구 방향을 논한다.

2. 관련 연구

사물인터넷 전용망(LTE Cat.M1)은 massive IoT 단말들의 초과 접속시 이를 해결하기 위한 기술이 필요하며, 이를 위해 기지국 파라미터인 inactivity timer 조절을 통해 자원을 관리함으로써 단말의 수용을 증가시킬 수 있다. 본 관련 연구에서는 inactivity timer에 대해 설명하고 무선 자원 관리를 위한 AI 관련 연구와 문제점을 서술한다.

2.1 Inactivity Timer 제어

Inactivity timer(T)는 단말의 데이터 전송이 완료된 시점부터 RRC 연결을 해제(Release)하기까지 기지국이 대기하는 시간이다[5]. 설정된 timer 값을 초과한 시간 동안 단말과 기지국 간 데이터 송수신이 없는 경우 RRC 접속을 해제하고, 다른 단말에게 RRC 자원을 제공한다.

기존 inactivity timer 관련 연구들은 주로 단말의 DRX(Discontinuous Reception) 메커니즘을 통한 에너지 효율 향상에 집중되어 왔다[6]. 그러나, 대규모 단말이 밀집된 massive IoT 환경에서는 단말 개별의 에너지 효율성보다 기지국의 한정된 무선 자원 고갈 문제를 선제적으로 방어하는 망 중심(Network-centric)의 관점이 요구된다.

Inactivity Timer를 길게 설정하는 경우 단말의 빈번한 재접속을 방지하여 시그널링 부하와 단말 전력 소모를 줄이지만, 기지국의 RRC 자원을 장시간 점유하여 신규 단말의 접속을 차단(Blocking)한다. 반면, Inactivity timer를 짧게하는 경우 자원을 회수하여 수용량을 높이지만, 제어 평면(Control Plane)의 부하를 가중시키고 단말의 수명을 단축시킨다. 또한, 비연속적 데이터를 전송하는 과정에서 전송 완료 전에 단말이 접속을 끊어 기지국 절단율이 상승하는 부작용이 있을 수 있다[7].

www.kci.go.kr

2.2 무선 자원 관리를 위한 AI 연구 및 한계

최근 급격히 변화하는 환경에 대처하기 위해 이동통신 환경에서도 AI를 활용한 기술들이 적용되고 있고, Deep Learning을 활용한 무선 자원 관리(RRM, Radio Resource Management) 연구가 활발히 진행되고 있다.

CNN 및 RNN의 기반의 예측 기술을 통해 과거 트래픽 패턴을 학습하여 미래 부하를 예측하고 자원을 할당하는 방식이 제안되었으나[8], IoT 단말의 비정기적(Sporadic) 트래픽 특성으로 인해 예측 정확도가 낮아 오탐율로 인한 문제가 발생한다.

심층 강화학습(DRL, Deep Reinforcement Learning)을 적용하는 경우 DQN(Deep Q-Network)을 이용해 스케줄링 우선순위나 전력 제어를 최적화하는 연구는 높은 성능을 보였다[9]. 그러나 대다수의 기존 연구는 (1) 보상 함수(Reward Function) 설계의 경직성으로 인해 자원 효율과 품질 사이의 동적인 가중치 변화를 반영하지 못하며, (2) 신경망의 높은 연산 복잡도로 인해 1ms 단위의 스케줄링 주기를 갖는 기지국 환경에서 실시간 추론 지연 문제를 야기한다[10].

본 논문은 이러한 기존 AI 연구의 한계를 극복하기 위해, AI가 직접 파라미터를 결정하는 대신 평형 방정식의 지능형 가중치를 최적화하고, 실제 제어는 경량화된 선형 수식으로 처리하는 하이브리드 구조를 제안한다.

3. AI 기반 Inactivity Timer 최적화 모델

사물인터넷 전용망(LTE Cat.M1)에서 inactivity timer 조절은 설정값에 따라 자원 가동률과 사용자 경험 간의 트레이드오프가 발생한다. 따라서 특정 상황에서만 유효한 고정 임계치 및 규칙 기반의 한계를 벗어나, 환경 변화에 맞추어 최적의 평형점을 실시간으로 이동시킬 수 있는 지능형 제어 구조가 필수적이다.

3.1 AI 기반 지능형 Inactivity Timer 최적화

본 논문에서 제안하는 지능형 평형 방정식은 물리적인 네트워크 자원 제약과 AI의 추상적인 보상 함수를 연결하는 교량 역할을 한다. 기지국의 잉여 자원이 부족해질수록 자원 회수의 시급성이 커지지만, 타이머를 과도하게 단축하면 호 절단율이 급증하여 서비스 품질이 훼손된다.

본 모델의 DQN 에이전트는 누적 보상 R을 극대화하는 방향으로 학습을 진행하며, 그 결로써 실시간 제어 가중치 ω_1, ω_2를 도출한다. 이 과정에서 가중치 ω_1은 보상 R을 구성하는 '접속 성공률'을 높이기 위해 기지국이 취할 수 있는 자원 회수의 한계 이익(Marginal Benefit, MB)을 조율하는 계수로 작용한다. 반대로 ω_2는 보상 R의 '절단율 페널티'를 억제하기 위해 감수해야 할 품질 열화의 한계 비용(Marginal Cost, MC)을 제어하는 계수로 작용한다.

에이전트가 전체 누적 보상 R을 극대화하는 최적점은 추가적인 자원 회수 이익과 품질 열화 비용이 상쇄되는 $MB = MC$ 상태이므로, 이를 물리적 파라미터와 결합하여 다음과 같은 지능형 평형 방정식을 연역적으로 유도할 수 있다.

$$\omega_1 \cdot \frac{x}{N} = \omega_2 \cdot \frac{y_2}{y_1} \tag{1}$$

여기서 N은 특정 기지국의 최대 RRC 접속 허용 단말 수, x는 현재 접속 중인 단말 수이며, y_1은 inactivity timer 변경 전 절단율, y_2는 inactivity timer 변경 후 절단율이다.

타이머 변화량(ΔT_I)에 따른 절단율의 변화를 큐잉 이론의 천이 확률에 기반하여 $y_2 \approx y_1 - k\Delta T_I$로 근사하고 [11], 식 (1)에 대입하여 ΔT_I에 대해 이항하면 실시간 실행 계층에서 사용할 최종 제어식을 얻는다.

$$\Delta T_I = \frac{y_1}{k} \cdot \left(1 - \frac{\omega_1}{\omega_2} \cdot \frac{x}{N}\right) \tag{2}$$

여기서 k는 타이머 변화에 따른 절단율 민감도를 나타내는 선형화 계수이다. 이는 하위 계층의 통신 물리적 특성을 반영하는 값으로, 상위 AI의 제어 변수 중복(Over-parameterization)을 방지하고 연산 복잡도를 줄이기 위해 기지국이 과거 트래픽 로그를 바탕으로 산출하여 보유하는 환경 상수로 취급한다. 반면, 가중치 ω_1, ω_2는 AI가 현재 네트워크의 상태를 분석하여 동적으로 도출하는 지능형 파라미터로, 자원 회수의 시급성과 서비스 품질 유지의 필요성을 실시간으로 대변한다. 결과적으로 본 모델은 AI가 도출한 최적 가중치를 입력받아 단순 사칙연산만으로 최적의 ΔT_I를 산출함으로써, 기지국의 연산 오버헤드를 최소화하고 실시간 스케줄링 주기 내에 제어 처리를 가능하게 한다.

www.kci.go.kr

3.2 제어 방향성의 수학적 일관성 확인

부하량 x가 상승할 때 제어 로직이 선제적 자원 확보로 이어지는지 검증하기 위해 식 (2)를 부하량 편미분하면 다음과 같다.

$$\frac{\partial \Delta T_I}{\partial x} = -\frac{\omega_1}{\omega_2} \cdot \frac{y_1}{k \cdot N} < 0 \quad (3)$$

이는 기지국 부하(x)가 상승할수록 타이머 변화량(ΔT_I) 가변치가 음의 방향으로 작동하여 실질적인 타이머 단축을 유발하며, 타이머 단축 방향과 부하 상승 방향이 설계 의도대로 일치함을 수식적으로 확인해 준다.

4. AI 하이브리드 모델링 및 실시간성 분석

본 장에서는 제안하는 지능형 평형 모델의 핵심인 DQN(Deep Q-Network)의 설계 구조와, 이것이 실제 기지국 스케줄링 환경에서 어떻게 실시간성을 보장하는지 분석한다.

4.1 DQN 기반 지능형 가중치 최적화 모델

기본적인 적응형 방식이 고정된 rule-based를 따르는 것과 달리, 본 모델은 네트워크 다변수 상태를 입력받아 가중치 ω_1, ω_2를 동적으로 도출한다.

1) 상태 공간 (State Space, S): 기지국의 물리적·논리적 상태를 반영하는 입력 파라미터로 현재 부하율(x/N), 무선 환경 지표(Avg. SNR), 트래픽 변화율(Δx), 누적 절단율을 사용한다.
2) 작용 공간 (Action Space, A): 에이전트가 하위 평형 방정식을 제어하기 위해 출력하는 최적 파라미터 조합인 (ω_1, ω_2) 세트를 의미한다. 하위 실행 계층은 AI가 도출한 이 지능형 가중치 세트와 기지국의 물리적 환경 상수 k를 결합하여 매 1ms마다 즉각적인 선형 제어를 수행한다. 이산화 매핑 시 시스템의 급격한 오실레이션을 방지하기 위해 이전 가중치 대비 ±5% 범위 내에서 미세 조정을 수행하도록 제한하였다.
3) 보상 함수 (Reward Function, R): 에이전트의 최종 목적 함수인 보상 R은 특정 가중치에 편향되지 않도록 객관적 환경 품질 지표의 합으로 다음과 같이 정의한다[12, 13].

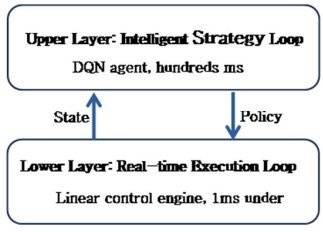
R = (Success Rate) - (Drop Rate Penalty) - (Signaling Penalty)

이때 서로 다른 물리적 단위(Scale)를 가진 각 지표들이 학습에 미치는 지배적 편향(Dominance)을 방지하기 위해, 모든 항목은 0과 1 사이의 값으로 최소-최대 정규화(Min-Max Normalization)를 거쳐 결합된다. DQN 에이전트는 이 객관적인 네트워크 성능 지표(R)를 극대화하는 방향으로 상태(State)를 학습한다. 학습의 결과로써 매 100ms 주기마다 최적의 제어 가중치인 작용 공간 (ω_1, ω_2)를 도출하여 하위 계층의 평형 방정식에 전달하는 전략-실행 분리(Strategy-Execution Separation) 구조를 채택하였다.

4.2 Hybrid 제어 아키텍처: AI-Strategy & Formula-Execution

본 논문이 제안하는 가장 큰 기술적 특징은 [Fig.1]과 같이 AI가 직접 타이머 값을 결정하지 않은 '전략-실행 분리'구조에 있다.

1) 상위 계층 (AI 전략 루프): 수백 ms 주기로 동작하며, DQN 에이전트가 기지국 환경(부하율, 평균 SNR, 절단율)을 상태 공간(State)으로 인지한다. 에이전트는 지능형 가중치 세트 (ω_1, ω_2)를 추론하고 하달한다.
2) 하위 계층 (실시간 실행 루프): 업데이트된 가중치를 기반으로, 식 (2)를 통해 매 RRC 연결 요청 시점마다 즉각적인 최적의 타이머 값을 산출하고 적용한다.



[Fig. 1] Strategy-Execution Separation Architecture Diagram

4.3 계층적 제어 구조의 연산 복잡도 분석

기존의 심층 강화학습 기반 직접 제어 모델은 기지국의 스케줄링 단위인 매 1ms(TTI)마다 상태 관측, 정규화, 그리고 신경망 파라미터에 대한 부동소수점 연산(FLOPs)

www.kci.go.kr

을 하위 제어 계층에서 강제적으로 수행해야 한다. 이로 인해 상용 임베디드 기지국 프로세서 환경을 가정한 프로파일링 분석에서, 신경망 연산 및 통신 오버헤드를 포함하여 평균 약 18.5ms의 추론 지연을 유발하며, 이는 실시간 RRC 자원 관리에 결함을 유발할 수 있다.

반면, 제안하는 Hybrid 모델은 '전략-실행 분리' 아키텍처를 통해 연산 복잡도를 낮춘다. 제안 모델은 연산량이 큰 DQN 가중치 추론(18.5ms)은 100ms 주기로 상위 계층에서 백그라운드로 분산 처리하고, 하위 계층(1ms 주기)에서는 단순 선형 1차 방정식만을 계산하므로 0.1ms 미만의 지연 시간 내에 즉시 실행된다. 결과적으로 무거운 신경망 갱신 오버헤드를 100회의 하위 실행 주기 동안의 시간 가중 평균으로 환산하면, 전체 시스템의 TTI 당 평균 제어 시간은 약 0.285ms ((18.5ms+(0.1ms×100))/100)로 수렴하게 된다. 이는 하드웨어의 추가적인 증설 없이도 3GPP에서 규정하는 RRC 재구성 주기 표준을 충족하며[14], 실제 물리 계층 스케줄링에 영향을 주지 않음을 확인할 수 있다.

4.4 모델의 Convergence 및 상용망 적용 안정성

강화학습을 실시간 상용 통신망에 직접 적용할 경우, 에이전트의 초기 무작위 탐색 과정에서 기지국 파라미터가 극단적으로 설정되어 대규모 통신장애(Signaling Storm)를 유발할 위험이 있다. 이러한 강화학습 초기 탐색 과정에서의 통신망 불안정을 방지하기 위해 제안 모델은 과거의 대규모 트래픽 이력 데이터를 활용하여 10,000 에피소드 이상의 오프라인 사전 학습을 진행함으로써 모델의 수렴성을 확보하였다. 실제 운영 단계에서는 무작위 탐색률(ε)을 0.01 이하로 제한하여 실시간 미세 조정만을 수행하도록 설계하여 통신망의 안전성을 보장한다.

5. 성능 평가 및 분석

본 장에서는 제안하는 모델의 성능을 검증하기 위해 NS-3(Network Simulator 3)의 LTE-EPC 네트워크 시뮬레이션 모듈(LENA)을 기반으로 실험 환경을 구축하고 비교 시뮬레이션을 수행하였다[15].

5.1 시뮬레이션 환경 및 파라미터 설정

Massive IoT 단말의 정적 특성을 반영하여 이동성 모델은 ConstantPositionMobilityModel을 적용하였고, 무선 채널 환경은 경로 손실과 매크로 셀 환경을 모사하기 위해 FriisPropagationLossModel을 사용하였다. 사물인터넷 전용망인 LTE Cat.M1의 자원 병목 현상을 모사하기 위해 상·하향 링크의 시스템 대역폭을 6 PRBs (1.4 MHz)로 제한하였다. 트래픽은 단말들이 비동기적으로 간헐적 데이터를 전송하는 특성을 반영하여 OnOffApplication을 커스텀하였으며, 패킷 크기 100~500 Bytes, 트래픽 발생률 λ=5~50 수준의 포아송 분포 발생기를 구현하여 적용하였다.

기지국의 최대 수용 한계(M)는 500개 단말로 제한하였으며, 단말(UE)의 접속 시도를 최대 1,200개까지 점진적으로 증가시키며 네트워크 과부하 상황을 유도하였다. 타이머 구현을 위해 MAC 계층의 버퍼 상태를 모니터링하여 데이터 송수신이 없을 경우 RRC 연결을 강제 해제(Connection Release)하는 'Custom Data Inactivity Timer'를 ns3::LteEnbRrc 클래스 내에 확장 구현하였다. 제안하는 '전략-실행 분리' 아키텍처를 위해 상위 계층(Python)에서 산출된 가중치는 ns3-gym 프레임워크를 통해 하위 계층(C++)으로 하달되어 이 커스텀 타이머의 평형 방정식을 동적으로 갱신하도록 스케줄링하였다. 통계적 신뢰성을 확보하기 위해 독립적인 난수 시드를 적용하여 총 100회의 몬테카를로 반복 시뮬레이션을 수행하였으며, 제시된 수치는 95% 신뢰구간 내의 평균값이다.

상세한 네트워크 및 AI 하이퍼파라미터 설정은 〈Table 1〉에서 설명한다.

〈Table 1〉 Simulation Environment and System Parameters

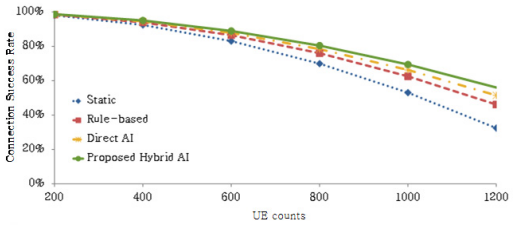
Category	Parameter	Value
Network Environment	Simulator	NS-3 (LTE LENA Module)
	Cell Radius	1km
	System Bandwidth	6 PRB (1.4MHz)
	Max RRC Capacity Limit (M)	500 UEs
	Total Number of UEs (x)	200 ~ 1,200 UEs
	Traffic Generation Model	Poisson, λ=5~50
	Packet Size	100~500 Bytes
Timer Settings	Default Inactivity Timer	10s
	Control Interval (Linearized Eq.)	1ms(TTI)
AI Model (DQN)	Agent Update Interval	100ms
	Learning Rate	0.001
	Discount Factor (γ)	0.95
	Replay Buffer Size	10,000

본 논문에서는 제안 모델의 성능을 다각도로 검증하기

위해 고정 (Static, 10초), 규칙 기반 적응형 (Rule-based Adaptive), AI 직접 제어 (Direct AI) 세 가지의 기존 제어 방식과 비교 분석을 수행하였다.

5.2 실험 결과 및 상세 분석

시뮬레이션은 접속 성공률, 최대 수용량 향상율, 그리고 절단율 측면에서의 생존성을 중점적으로 비교 분석하였다.



[Fig. 2] Connection Success Rate

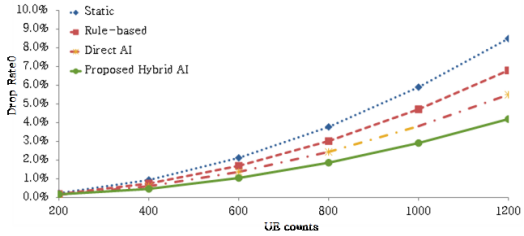
접속 성공률은 단말 접속 시도가 기지국 수용 한계인 500 UEs를 초과하는 과부하 구간부터 제어 방식 간의 성능 차이가 나타났다. 극한 과부하 상황인 단말 수가 1,200 UEs에 도달했을 때, 고정 방식은 자원 회수 지연으로 인해 접속 성공률이 33%까지 급락하여 심각한 통신 장애를 유발했다. 규칙 기반 적응형 방식 역시 46%의 성공률에 그치며 트래픽 폭증에 유연하게 대응하지 못했다. 반면, 제안하는 하이브리드 모델은 동일한 1,200 UEs 환경에서 56%의 접속 성공률을 기록하였다[Fig.2]. 이는 부하율(x/N)이 상승하는 초기 단계부터 AI가 가중치 ω_1을 미세하게 상향시켜, 선형 수식을 통해 선제적이고 연속적으로 자원을 확보한 결과로 해석된다.

제안하는 하이브리드 제어 모델은 $\partial \Delta T_{\#}/\partial x < 0$ 원리에 따라 부하 상승 시 타이머를 선제적으로 단축하여 자원 회전율을 높인다. 그 결과, 추가 장비 증설 없이도 소프트웨어 최적화만으로 유효 수용 단말 수가 기존 대비 약 24% 향상되었다. 이는 〈Table 2〉에서 보이는 바와 같이 접속 성공률 70% 기준 유효 수용 단말 수를 800대에서 992대로 확장한 수치이다. 여기에서 유효 수용 단말 수는 기지국의 '동시 접속 수용 한계(N=500)'와 달리 접속 성공률 70% 임계치를 만족하면서 셀 내에서 서비스 가능한 '전체 트래픽 발생 단말 수'를 의미한다.

〈Table 2〉 Comparison of Maximum Capacity Enhancement Index (Based on Connection Success Rate ≥ 70%)

Control Scheme	Effective Capacity (No. of UEs)	Capacity Enhancement Index
Static (Baseline)	800 UEs	100% (Baseline)
Rule-based Adaptive	896 UEs	112%
Conventional Direct AI	944 UEs	118%
Proposed Hybrid AI	992 UEs	124%

Inactivity timer를 단축하면 자원은 확보되나 단말의 잦은 재접속으로 인해 호 절단율이 상승하는 부작용이 있다. 실험 결과, [Fig.3]에서 보이는바와 같이 극한 과부하(1,200 UEs) 환경에서 AI 직접 제어 방식은 자원 확보에 치중한 나머지 절단율이 5.5%로 나타났고, 고정 방식(Static)은 최대 8.5%의 절단율을 기록했다.



[Fig. 3] Drop Rate

고정 방식에서 절단율이 최대 8.5%까지 치솟은 원인은 긴 타이머로 인한 자원 미회수로 접속이 차단된 단말들이 공격적인 재접속 시도를 반복하면서 시그널링 폭풍(Signaling Storm)이 유발되었고, 이로 인해 기지국의 제어 평면자원 자체가 붕괴하여 기존 연결마저 비정상적으로 절단된 결과로 해석된다. 반면, 제안 모델은 평형 방정식 내에 품질 방어 가중치(ω_2)를 독립적으로 두어 이 문제를 완화하였다. AI는 단말의 평균 SNR이 저하되거나 핑퐁 접속이 빈번해지는 패턴을 인지하면 가중치 ω_2를 높여 과도한 타이머 단축을 억제하였다. 그 결과, 과부하 상태에서도 절단율을 4.2% 수준으로 유지하며 서비스 연속성 측면에서 안정성을 만족한다[16].

www.kci.go.kr

6. 결론 및 향후 연구 방향

본 논문은 대규모 IoT 기기가 밀집된 사물인터넷 전용망(LTE Cat.M1) 환경에서 발생하는 기지국의 RRC 자원 고갈 및 제어 평면 과부하 문제를 해결하기 위해, ‘AI 기반 지능형 하이브리드 제어 모델’을 제안하였다. 기존의 고정형 inactivity timer 방식이나 단순 임계치 기반의 적응형 기법이 가지는 경직성을 극복하기 위해, 기지국의 실시간 부하율(x/N)과 서비스 품질(절단율) 간의 관계를 수식화한 지능형 평형 방정식을 고안하였다.

본 연구의 기술적 성과는 AI의 무거운 연산과 기지국의 엄격한 스케줄링 요구사항 사이의 딜레마를 ‘전략과 실행의 분리’를 통해 해결한 점이다. 결론적으로, 본 제안 모델은 지능적 상황 판단 능력과 수학적 실행의 즉각성을 결합함으로써 massive IoT 환경의 통신 자원 관리 능력을 끌어올렸다. 이는 소프트웨어 고도화만으로 접속 밀도 한계를 경제적으로 돌파할 수 있음을 의미하며, 향후 5G-Advanced 및 6G 초연결 네트워크의 자원 할당 최적화를 위한 기반 기술로 활용될 수 있을 것으로 기대된다.

REFERENCES

[1] 3GPP, "TS 36.300: E-UTRA and E-UTRAN; Overall description; Stage 2," V15.0.0, 2018.

[2] 3GPP, "Technical Specification Group Radio Access Network; NR; User Equipment (UE) radio access capabilities (Release 17)," 3GPP TS 38.306, V17.4.0, 2023.

[3] Y.P.E.Wang, X.Lin, A.Adhikary, A.Grovlen, Y.Sui, Y.Blankenship, J.Bergman, and H.S.Hazra, "A Primer on 3GPP Narrowband Internet of Things," IEEE Communications Magazine, Vol. 55, No. 3, pp. 117-123, 2017.

[4] R.Ratasuk, N.Mangalvedhe, Y.Zhang, M.Robert, and J.Koskinen, "Overview of narrowband IoT in LTE Rel-13," in Proceedings of the 2016 IEEE Conference on Standards for Communications and Networking (CSCN), Berlin, Germany, pp. 1-7, 2016.

[5] Y.Yamada and Y.Miya, "Impact of inactivity timer on performance of control plane in LTE core network under burst traffic," in Proceedings of the 2015 IEEE Conference on Communications (ICC), London, UK, pp. 4668-4673, 2015.

[6] K.Lin, H.Liu, K.Hu, A.Huang, and H.Wei, "A Survey on DRX Mechanism: Device Power Saving from LTE and 5G New Radio to 6G Communication Systems," IEEE Communications Survey & Tutorials, Vol. 25, No. 1, pp. 156-183, 2023.

[7] Z.Zhang, Z.Zhao, H.Guan, D.Miao, and Z.Tan, "Study of Signaling Overhead Caused by Keep-Alive Messages in LTE Network," in Proceedings of the IEEE 78th Vehicular Technology Conference (VTC Fall), Las Vegas, NV, USA, pp. 1-5, 2013.

[8] H.Sun, X.Chen, Q.Shi, M.Hong, X.Fu, and N.D. Sidiropoulos, "Learning to Optimize: Training Deep Neural Networks for Interference Management," IEEE Transactions on Signal Processing, Vol. 66, No. 20, pp. 5438-5453, 2018.

[9] N.C.Luong, D.T.Hoang, S.Gong, D.Niyato, P.Wang, Y.C. Liang, and D.I.Kim, "Applications of Deep Reinforcement Learning in Communications and Networking: A Survey," IEEE Communications Surveys & Tutorials, Vol. 21, No. 4, pp. 3133-3174, 2019.

[10] J.Chen and X.Ran, "Deep Learning With Edge Computing: A Review," Proceedings of the IEEE, Vol. 107, No. 8, pp. 1655-1674, 2019.

[11] L.Kleinrock, Queueing Systems, Volume 1:Theory, Wiley, 1975.

[12] V.Mnih, K.Kavukcuoglu, D.Silver, A.A.Rusu, J.Veness, M.G.Bellemare, A.Graves, M.Riedmiller, A.K.Fidjeland, G.Ostrovski, S.Petersen, C.Beattie, A.Sadik, I.Antonoglou, H.King, D.Kumaran, D.Wierstra, S.Legg, and D.Hassabis, "Human-level control through deep reinforcement learning," Nature, Vol. 518, No. 7540, pp. 529-533, 2015.

[13] F.B.Mismar, B.L.Evans, and A.Alkhateeb, "Deep Reinforcement Learning for 5G Networks: Joint Beamforming, Power Control, and Interference Coordination," IEEE Transactions on Communications, Vol. 68, No. 3, pp. 1581-1592, 2020.

[14] 3GPP, "Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification (Release 15)," 3GPP TS 36.331, V15.4.0, 2019.

[15] T.R.Henderson, M.Lacage, G.F.Riley, C.Dowell, and J.Kopena, "Network Simulations with ns-3," ACM SIGCOMM Educational Workshop, Vol. 15, No. 4, pp. 78-84, 2008.

[16] W.Zhan, C.Xu, X.Sun, and J.Zou, "Toward Optimal Connection Management for Massive Machine-Type Communications in 5G System," IEEE Internet of Things Journal, Vol. 8, No. 17, pp. 13237-13250, 2021.

www.kci.go.kr

박 복 녕(Bok-Nyong Park) [정회원]



■ 2008년 2월 : 고려대학교 컴퓨터학과 (이학박사)
■ 2007년 8월 ~ 현재 : LG유플러스 NW부문 (팀장)

〈관심분야〉
사물인터넷, 차세대 이동통신, AI, Autonomous Network

권 준 혁(June-Hyuk Kwon) [정회원]



■ 1991년 : 연세대학교 경영학과 졸업
■ 1993년 : 한국과학기술원 경영과학과 (공학석사)
■ 1998년 : 한국과학기술원 경영공학과 (공학박사)
■ 1998년 ~ 현재 : LG유플러스 NW부문 (부사장)

〈관심분야〉
차세대 유무선통신, Autonomous Network, AI

www.kci.go.kr