

B5G 네트워크에서의 네트워크 데이터 분석을 위한 트래픽 효율적인 데이터 집계 기법

김보경, 고한얼
경희대학교

bokyeong0405@khu.ac.kr, heko@khu.ac.kr

Traffic-Efficient Data Aggregation Scheme for Network Data Analytics in B5G Networks

Kim Bokyeong, Ko Haneul
Kyung Hee Univ.

요 약

네트워크 데이터 분석 기능(NWDAF)에서 정확한 분석을 도출하려면 다양한 소스로부터 많은 데이터를 수집해야 하는데, 이로 인해 발생할 수 있는 네트워크 혼잡을 해결하기 위해 데이터를 생성될 때마다 전달하는 대신 데이터를 집계할 수 있다. 그러나 데이터를 더 많이 집계할수록 NWDAF는 더 지연된 정보를 갖게 된다. 네트워크 혼잡과 데이터 신선도 간의 균형을 맞추기 위해, 본 논문에서는 트래픽 효율적인 데이터 집계 방안(TE-DAS)을 제안하고, 데이터 집계 정책을 도출한다. 구체적으로는 데이터의 신선도를 유지하면서 일시적인 네트워크 혼잡을 최소화하는 것을 목표로 하며, 이를 제약된 마르코프 결정 문제(CMDP)로 공식화한다. 성능평가 결과는 도출된 데이터 집계 정책이 충분한 데이터 신선도를 제공하면서 네트워크 혼잡을 90% 이상 줄일 수 있음을 보여준다.

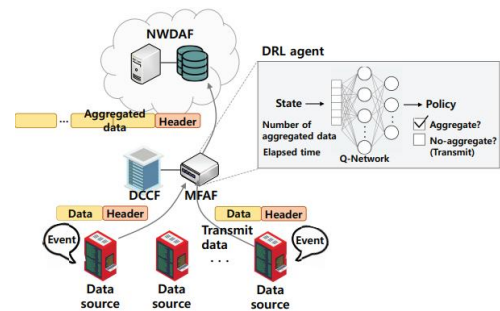
I. 서 론

셀룰러 네트워크에서의 최적화와 자동화를 위해 기계 학습(ML)과 인공지능(AI) 모델의 활용이 중요하다[1]. 이러한 맥락에서 3GPP는 네트워크 데이터 분석 기능(NWDAF)을 표준 기능으로 도입하여 네트워크 데이터 분석을 강화했다[2]. NWDAF는 ML/AI 알고리즘을 사용하여 분석을 도출하고 이러한 분석을 다른 네트워크 기능(NF)에 제공한다. NWDAF에서 정확한 분석을 도출하려면 다양한 소스(예: 네트워크 기능(NF))로부터 많은 데이터를 수집해야 하는데, 이는 네트워크 혼잡을 초래할 수 있다. 이러한 혼잡을 완화하기 위해 데이터가 생성될 때마다 전달하는 대신 데이터를 집계할 수 있지만, 더 많은 데이터를 집계할수록 NWDAF는 더 지연된 정보(즉, 데이터 신선도가 보장되지 않음)를 갖게 된다. 네트워크 혼잡과 데이터 신선도 간의 균형을 맞추기 위해, 본 논문에서는 트래픽 효율적인 데이터 집계 방안(TE-DAS)을 제안하고 데이터 집계 정책을 도출한다. 구체적으로, 데이터 집계 문제를 데이터 신선도를 일정 수준 이상 유지하면서 일시적인 네트워크 혼잡을 최소화하는 것을 목표로 한 제약된 마르코프 결정 문제(CMDP)로 공식화한다. 성능평가 결과는 도출된 데이터 집계 정책이 데이터 신선도를 충분히 제공하면서 비집계 정책에 비해 일시적인 네트워크 혼잡을 90% 이상 줄일 수 있음을 보여준다.

II. 시스템 모델

[그림 1]에서는 이벤트가 발생했을 때 NF가 데이터를 MFAF로 전달하면 이를 MFAF가 집계해 NWDAF로

전달하는 과정을 설명한다. 기본적으로 MFAF는 데이터를 NWDAF로 전달하기 전에 데이터를 형식화하고 전처리한다[3]. 특정 이벤트(즉, 데이터)가 자주 발생하는 상황에서 MFAF는 혼잡을 완화하기 위해 데이터를 일정 시간 동안 집계한 후 NWDAF로 전달할 수 있다. 이 때 네트워크 혼잡과 데이터 신선도를 고려한다. 데이터를 집계할지 여부는 집계된 데이터의 양과 경과 시간에 따라 결정된다. MFAF에서는 CMDP 모델과 강화 학습을 사용하여 최적의 데이터 집계 정책을 찾는다.



[그림 1] 시스템 모델

III. 강화 학습 문제

데이터 생성 패턴에 대한 사전 지식 없이 최적의 데이터 집계 정책을 얻기 위해 강화 학습 접근 방식을 도입한다.

A. 상태 공간

전체 상태 공간 S 는 다음과 같이 정의될 수 있다.

$$S = N \times E$$

여기서 N 은 MFAF 에서 집계된 데이터의 개수에 대한 상태 공간을 나타내고, E 는 첫 집계부터 경과된 시간에 대한 상태 공간을 나타낸다.

B. 행동 공간

행동 공간 A 는 다음과 같이 정의될 수 있다.

$$A = \{0, 1\}$$

여기서 A 는 에이전트(즉, MFAF)가 데이터를 집계하는지 여부를 나타낸다. 즉, 에이전트는 $A = 0$ 이면 데이터를 집계하고, $A = 1$ 이면 집계하지 않고 데이터를 NWDAF 로 전송한다.

C. 전이 확률

에이전트가 집계된 데이터를 전송할 때(즉, $A = 1$) N 은 0이 되므로 N 의 전이는 A 의 영향을 받는다. 한편, E 는 집계된 데이터가 있을 때만 증가한다. 또한, 에이전트가 집계된 데이터를 전송할 때 E 는 리셋된다. 즉, E 에 대한 상태 전이는 N 과 A 의 영향을 받는다. 따라서 현재 상태 $S = [N, E]$ 에서 다음 상태 $S' = [N', E']$ 로의 전이 확률은 다음과 같이 설명될 수 있다.

$$P[S'|S, A] = P[N'|N, A] \times P[E'|E, N, A]$$

$P[N'|N, A]$ 와 $P[E'|E, N, A]$ 에 대한 구체적인 정의는 분량 제한으로 생략한다.

D. 비용 함수

네트워크 혼잡을 최소화하기 위해 비용 함수 $r(S, A)$ 를 정의한다. 에이전트가 전송 행동(즉, $A = 1$)을 선택하면 네트워크 혼잡이 발생할 수 있으며, 이는 패킷 크기에 비례한다. 패킷 헤더와 단일 데이터의 크기를 각각 T_H 와 T_D 라고 하면, 패킷 크기는 $T_H + NT_D$ 로 계산될 수 있다. 따라서 $r(S, A)$ 를 다음과 같이 표현할 수 있다.

$$r(S, A) = \begin{cases} 0, & \text{if } A = 0 \\ T_H + NT_D, & \text{if } A = 1 \end{cases}$$

E. 제약 함수

E 를 일정 수준 이하로 유지하기 위해 다음과 같이 제약 함수 $c(S, A)$ 를 정의한다.

$$c(S, A) = E$$

F. 최적화 공식

CMDP 모델은 아래와 같이 정의할 수 있다.

$$\begin{aligned} & \min_{\pi} \zeta_C \\ & \text{s. t. } \Psi_E \leq \theta_E \end{aligned}$$

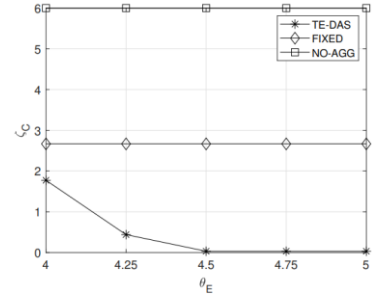
여기서 π 는 각 상태에서 특정 행동을 선택할 확률을 나타내는 정책을 의미하고, θ_E 는 E 의 상한값을 나타낸다. ζ_C 와 Ψ_E 는 각각 평균 네트워크 혼잡 정도와 첫 집계부터 경과된 시간의 평균으로 구체적인 수식은 분량 제한으로 생략한다. CMDP 모델에 대한 최적의 정책을 얻기 위해 위의 모델을 LP 문제로 변환하여 LP solver(예: Vaidya 알고리즘)를 이용하여 문제를 해결할 수 있다.

IV. 성능평가

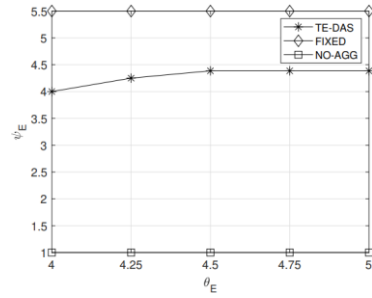
성능평가를 위해 TE-DAS 를 다음 두 가지 방식과 비교한다: 1) NO-AGG, 여기서 MFAF 는 데이터를 집계하지 않는다; 2) FIXED, 여기서 MFAF 는 집계된 데이터의 개수가 일정 수치(N_F)에 도달하면 데이터를 전송한다. 성능지표로는 ζ_C 와 Ψ_E 를 사용하며, N_F 는 3이다. 데이터 생성 속도는 0.3이며 패킷 헤더와 단일 데이터의 크기는 각각 1과 5이다.

[그림 2](a)와 (b)는 θ_E 가 ζ_C 와 Ψ_E 에 미치는 영향을 보여준다. TE-DAS 는 집계된 데이터의 양과 초기 집계 이후 경과된 시간을 고려하여 동적으로 데이터 집계할지 여부를 결정하여 Ψ_E 를 θ_E 이하로 유지하면서 ζ_C 를 최소화할 수 있다. 한편, 다른 방식들은 운영 환경을

고려하지 않은 고정된 정책을 따르기 때문에, θ_E 와 상관없이 성능이 일정하다.



(a) 평균 네트워크 혼잡 정도 ζ_C



(b) 첫 집계부터 경과된 시간의 평균 Ψ_E
[그림 2] θ_E 의 영향

V. 결론

본 논문에서는 트래픽 효율적인 데이터 집계 방안(TE-DAS)을 소개하고 데이터 집계 전략을 조사한다. 성능평가 결과는 제안된 데이터 집계 정책이 데이터 신선도를 유지하면서 일시적인 네트워크 혼잡을 크게 줄일 수 있음을 보여준다. 향후 연구에서는 강화 학습 접근 방식을 사용하여 환경에 대한 사전 지식 없이 최적의 집계 정책을 확립하는 방안을 더욱 발전시킬 계획이다.

ACKNOWLEDGMENT

본 연구는 2024 년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (과제번호: RS-2024-00340698)

참고 문헌

- [1] S. Sevçican *et al.*, "Intelligent Network Data Analytics Function in 5G Cellular Networks using Machine Learning," *Journal of Communications and Networks (JCN)*, vol. 22, no. 3, pp. 269–280, Jun. 2020.
- [2] 3GPP TS 23.501, "System architecture for the 5G System (5GS)," version 17.1.1, Jun. 2021.
- [3] 3GPP TS 23.288, "Architecture Enhancements for 5G System (5GS) to Support Network Data Analytics Services," version 17.2.0, Sep. 2021.