

디지털 트윈 응용을 위한 실시간 작업 스케줄링 방안

김천용 정민채

세종대학교 전자정보통신공학과

Real-time Task Scheduling for Digital Twin Applications

Cheonyong Kim

Minchae Jung

Department of Electronics and Information Engineering, Sejong University

cykim0807@sejong.ac.kr

mcjung@sejong.ac.kr

요 약

본 논문에서는 디지털 트윈 응용에서의 실시간 컴퓨팅 작업 스케줄링 성능을 평가하는 방안에 대해서 연구하였다. 특히, 물리 객체의 데이터를 이용해 디지털 객체의 상태 갱신을 위한 갱신 작업과 디지털 객체로부터 정보를 추출하는 추론 작업의 관계에 따라 스케줄링 정책을 수립하고, 이를 반영한 실시간 스케줄링 방안을 제안하였다. 제안하는 스케줄링 방안이 기존 방안에 비해 디지털 트윈 응용에 더 적합함을 실험을 통해 입증하였다.

1. 서론

디지털 트윈(DT, Digital Twin)은 물리 객체(PT, Physical Twin)의 실시간 상태를 반영하는 사이버 객체(CT, Cyber Twin)를 생성하고, 이를 활용해 최적화, 미래 상태 예측, 시뮬레이션을 수행할 수 있는 기술이다 [1]. 이를 통해 제조, 전력, 의료, 교통 등 다양한 산업 분야에서 효율성을 향상하고 비용을 절감할 수 있을 것으로 기대된다 [2].

DT 응용에서는 두 가지의 컴퓨팅 작업이 활용된다. 첫째, 갱신작업(update task)은 PT에서 수집된 센서 데이터를 이용해 CT를 갱신함으로써 PT와 CT를 동기화한다. 둘째, 추론작업(inference task)은 CT를 활용한 시뮬레이션을 통해 PT에 대한 다양한 정보를 도출하기 위해 사용된다. 실시간 작업 처리의 관점에서, 갱신작업과 추론작업은 독립적인 실시간 요구사항을 갖는다. 즉, 갱신작업은 PT와 CT간 동기화를 위해 저지연으로 처리되어야 하며, 추론작업은 추론요청에 대한 실시간 응답을 위해 즉각적으로 처리되어야 한다. 이러한 상충하는 실시간 요구사항으로 인해 DT 응용에 대한 실시간 작업 스케줄링 방안에 대한 연구가 필요하다. 기존 연구에서 CT 갱신 신뢰성 [3], DT 서비스 혼잡제어 [4], CT 마이그레이션 [5], 연합학습 기반 DT [6] 등 DT 응용과 관련된 다양한 문제가 연구되었지만, 실시간 작업 스케줄링에 대한 연구는 이루어지지 않았다.

본 논문에서는 DT freshness 기반 실시간 작업 스케줄링 방안을 제안한다. DT freshness는 갱신작업과 추론작업의 순서에 따라 정의되며, 갱신작업의 동기화 성능과 추론작업의 정확도를 포괄적으로 평가할 수 있는 평가지표이다. 제안방안은 제한시간 기반 작업 재정렬 기법을 활용해 작업 스케줄링을 수행하며, 이를 통해 많은 PT 및 CT를 포함하는 DT 응용에서 높은 DT freshness를 달성할 수 있다.

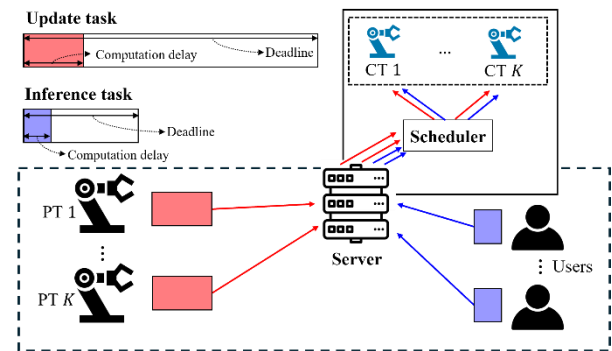


그림 1. DT 응용의 실시간 작업 스케줄링

2. DT Freshness 기반 실시간 스케줄링 방안

본 논문에서는 그림 1과 같이 다수의 PT와 CT, 그리고 사용자를 포함하는 DT 응용을 고려한다. 각 PT는 CT와의 동기화를 위해 주기적으로 갱신작업을 생성하며, 사용자로부터의 추론작업은 임의로 생성된다. 각 컴퓨팅 작업은 도착시간, 계산지연, 그리고 제한시간을 갖는다. 먼저 갱신작업은 상대적으로 긴 계산지연과 제한시간을 가지며, 주기적으로 생성된다는 점을 고려하여 갱신작업의 생성주기와 제한시간이 같다고 가정하였다. 반면, 추론작업의 경우, 상대적으로 짧은 계산지연과 제한시간을 갖는다. 반복적으로 수행되는 갱신작업을 고려하여, 본 논문에서는 한 주기 동안의 실시간 스케줄링을 다룬다.

이러한 환경에서, DT freshness는 갱신작업과 추론작업의 처리순서에 의해 정의된다. 그림 2는 PT의 갱신작업을 처리함에 따라 변화하는 CT의 상태변화와 이 상태변화 이전과 이후에 처리되는 추론작업들을 보여준다. 그림 2의 (a)와 같이 갱신작업 이전에 처리되는 추론작업은 PT의 최신정보를 반영하지 못한다. 만약 갱신작업이 PT의 중대한 상태변화를 포함하는 경우, 최신정보를 반영하지 못한 추론작업

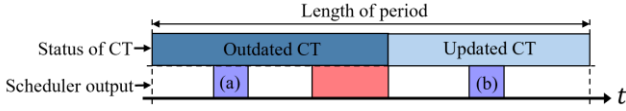


그림 2. 갱신작업에 의한 CT 상태 변화

의 결과는 DT 응용의 성능을 크게 저하할 수 있다. 따라서 DT 응용을 위한 실시간 스케줄링의 목적은 가능한 많은 추론작업이 그림 2의 (b)와 같이 갱신작업 이후에 처리되도록 하는 것이다. 이러한 관점에서, DT freshness의 정량적인 측정은 갱신작업 이후에 처리되는 추론작업의 수로 정의될 수 있다.

DT freshness를 향상하기 위해, 갱신작업과 추론작업의 관계를 분석함으로써 실시간 스케줄링 정책을 수립할 수 있다. 먼저, 갱신작업의 우선순위가 높을수록 높은 DT freshness를 달성할 수 있다. 예를 들어, 주기가 시작할 때 CT k 에 대한 갱신작업이 처리되는 경우, 이후 CT k 에 대한 모든 추론작업이 PT k 의 최신정보를 반영할 수 있다. 하지만, 모든 CT에 대한 갱신작업을 먼저 처리할 경우, 일부 추론작업의 제한시간이 위반될 수 있다. 따라서 각 추론작업의 제한시간을 고려하여 일부 갱신작업의 처리를 지연시켜야 한다. 이는 DT freshness를 다소 저하시키는 반면 추론작업의 처리를 보장할 수 있다.

이러한 갱신작업과 추론작업 간 관계를 고려하여, 제안방안은 제한시간 기반 작업 재정렬 기법을 이용한 실시간 스케줄링 방안을 제안한다. 제안방안은 크게 두 단계로 구성된다. 첫째는 제한시간 정렬 단계에서는 모든 추론작업들의 계산을 각자의 제한시간에 맞춰 정렬한다. 이때, 시간 축에서의 중복이 발생하지 않도록 함으로써 각 추론작업의 처리를 보장한다. 둘째는 작업 재정렬 단계에서는 갱신작업을 배치하며 작업들을 재정렬한다. 제한시간 정렬 단계의 결과에서, 추론작업이 점유하지 않은 시간을 확보할 수 있다. 즉, 주기의 앞부분부터, 확보된 시간이 갱신작업의 계산지연보다 큰 경우 갱신작업을 배치하고, 그렇지 않은 경우, 추론작업을 기존에 배치된 갱신작업 직후로 정렬한다. 이러한 두 단계의 작업 재정렬을 통해 프로세서의 이용률을 높이고 모든 작업의 처리를 보장할 수 있다. 게다가 추론작업의 제한시간을 보장하면서도 갱신작업을 먼저 처리함으로써 높은 DT freshness를 달성할 수 있다.

3. 실험 결과 및 결론

실험을 위해, K 개의 PT 및 CT를 포함하는 DT 응용을 고려했으며, 변수의 설정은 기존 연구에서 고려한 서버의 컴퓨팅 능력, 작업의 복잡도 및 빈도를 참고했다 [4][5]. 각 CT에 대한 갱신작업의 계산지연과 제한시간은 각각 0.1초와 10초이며, 추론작업의 계산지연과 제한시간은 각각 0.01초와 0.5초이다. 추론작업은 0.5초와 1초 사이의 간격을 두고 임의적으로 발생한다. 제안방안의 성능을 평가하기 위해 대표적인 실시간 작업 스케줄링 방안인 Earliest Deadline First (EDF)와 비교하였다. 그림은 K 의 증가

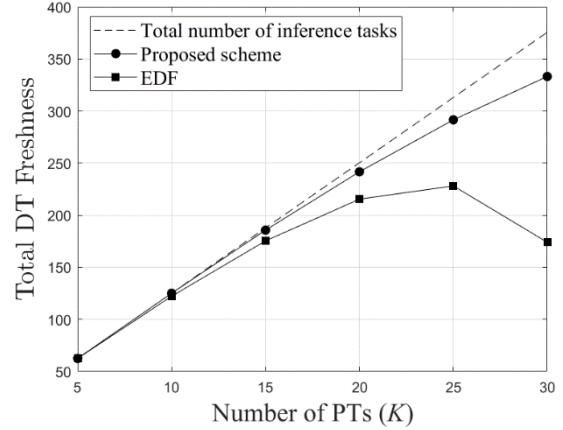


그림 3. PT 수에 따른 DT freshness 총합 결과

에 따른 DT freshness 총합 결과를 나타낸다. 그림에서 점선은 한 주기 동안 발생한 전체 추론작업의 수를 나타내며, DT freshness 총합의 상한을 의미한다. EDF는 제한시간이 짧은 작업이 먼저 수행되기 때문에, K 가 작을 때는 전체 추론작업의 수가 증가하면서 DT freshness 총합이 증가하지만, K 가 증가함에 따라 갱신작업이 지연되어 DT freshness 총합이 감소한다. 반면에 제안방안은 K 가 증가함에 따라 달성하는 DT freshness 총합도 증가한다. 다만, K 가 증가함에 따라, 주기의 앞부분에서 처리할 수 있는 갱신작업의 수가 제한적이기 때문에 DT freshness 상한과의 차이가 증가하지만, EDF에 비해 상한에 근접한 DT freshness 총합을 달성할 수 있다.

4. ACKNOWLEDGEMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2021R1C1C1012950).

5. 참고 문헌

- [1] O. Hashash, C. Chaccour, W. Saad, K. Sakaguchi and T. Yu, "Towards a Decentralized Metaverse: Synchronized Orchestration of Digital Twins and Sub-Metaverses," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2023, pp. 1905-1910.
- [2] L. U. Khan, Z. Han, W. Saad, E. Hossain, M. Guizani and C. S. Hong, "Digital Twin of Wireless Systems: Overview, Taxonomy, Challenges, and Opportunities," *IEEE Commun. Surv. Tutor.*, vol. 24, no. 4, pp. 2230-2254, Fourthquarter 2022.
- [3] O. Hashash, C. Chaccour, and W. Saad, "Edge continual learning for dynamic digital twins over wireless networks," in *Proc. IEEE Int. Workshop Signal Process. Advances Wireless Commun. (SPAWC)*, 2022, pp. 1-5.
- [4] X. Lin, J. Wu, J. Li, W. Yang, and M. Guizani, "Stochastic digital twin service demand with edge response: An incentive-based congestion control approach," *IEEE Trans. Mobile Comput.*, vol. 22, no. 4, pp. 2402-2416, 2023.
- [5] Y. Lu, S. Maharjan, and Y. Zhang, "Adaptive edge association for wireless digital twin networks in 6G," *IEEE Internet Things J.*, vol. 8, no. 22, pp. 16 219-16 230, 2021.
- [6] Y. Lu, X. Huang, K. Zhang, S. Maharjan, and Y. Zhang, "Low-latency federated learning and blockchain for edge association in digital twin empowered 6G networks," *IEEE Trans. Ind. Informat.*, vol. 17, no. 7, pp. 5098-5107, 2021.