

## Task Scheduling Scheme for Maintaining Freshness in Digital Twin Systems

Taehun Yang\*, Cheonyong Kim\*\*

\*Professor, School of IT Convergence Engineering, Gyeongsang National University, Andong, Korea  
\*\*Professor, Dept. of Computer Science, Korea National University of Transportation, Chungju, Korea

### [Abstract]

In digital twin systems, maintaining a high level of freshness for cyber twins is essential for accurate inference and real-time decision making. However, update tasks and inference tasks are processed under limited computing resources, causing a trade-off between task throughput and digital twin freshness. Existing scheduling methods prioritize task throughput, resulting in degraded digital twin freshness under a high traffic arrival rate. To address this issue, we propose an expected-freshness-based scheduling scheme that dynamically determines the execution timing of update tasks by predicting the expected number of inference tasks within each update interval. By estimating the maximum achievable freshness and scheduling update tasks when degradation is expected, the proposed scheme ensures the required digital twin freshness level. Simulation results show that our scheme maintains digital twin freshness even under increasing task arrival rates, while incurring only moderate task rejection. The proposed scheduling scheme is suitable for applications where freshness is more critical than task throughput.

▶ **Key words:** Digital Twin System, Freshness, Real-time Scheduling, Update-Inference Trade-off, Task Prioritization

### [요 약]

디지털 트윈에서 응용의 품질을 향상하기 위해 디지털 객체에 대한 높은 신선도가 필요하다. 그러나 갱신 및 추론 작업은 제한된 컴퓨팅 자원을 공유하므로, 신선도와 작업 처리율 사이의 상충관계가 발생한다. 기존 연구는 작업 처리율을 우선하므로 작업 도착률이 증가할 때 신선도가 저하된다. 이러한 문제를 해결하기 위해 본 논문에서는 예산 신선도 기반 스케줄링 방안을 제안한다. 제안 방안은 실시간으로 예산 신선도를 계산하여 갱신 작업 수행 시점을 동적으로 조정함으로써 요구 신선도를 만족한다. 시뮬레이션 결과는 제안 방안이 일부 작업을 거부함으로써 요구되는 신선도를 유지할 수 있음을 보인다. 제안 방안은 작업 처리율보다 신선도 보장이 중요한 응용에 적합한 스케줄링 방안으로 활용될 수 있다.

▶ **주제어:** 디지털 트윈 시스템, 신선도, 실시간 스케줄링, 갱신-추론 상충관계, 작업 우선순위

• First Author: Taehun Yang, Corresponding Author: Cheonyong Kim  
\*Taehun Yang (thyang@gknu.ac.kr), School of IT Convergence Engineering, Gyeongsang National University  
\*\*Cheonyong Kim (cykim@ut.ac.kr), Dept. of Computer Science, Korea National University of Transportation  
• Received: 2026. 01. 12, Revised: 2026. 01. 23, Accepted: 2026. 02. 05.

## I. Introduction

사물인터넷(Internet of Things, IoT)은 대규모의 실시간 데이터 수집을 통해 제조, 헬스케어, 교통, 스마트시티 등 다양한 산업 분야의 서비스 품질을 향상하는 기반을 제공한다. 이러한 IoT 데이터를 활용하면 디지털 트윈(Digital Twin: DT) 개념을 통해 물리적 시스템을 가상 공간에 재현할 수 있다[1-2]. DT는 현실의 물리적 객체(PT)와 동일한 특성, 기능, 동작을 수행하는 대응 객체인 사이버 트윈(CT)을 구성하여, 이 두 객체가 실시간 데이터 흐름을 통해 상호 진화하는 시스템이다. 결과적으로, DT는 물리적 시스템의 효율을 향상시키고 이상 감지 및 예측 제어를 가능하게 하는 핵심 기술로 자리 잡고 있다.

DT 응용 분야는 크게 (1) 실시간 모니터링 및 최적화, (2) 테스트 및 시뮬레이션으로 나눌 수 있다. 첫째, 실시간 모니터링 및 최적화의 경우, CT는 대응하는 PT의 상태를 실시간으로 반영하며, 센서 데이터 및 내부 로그를 포함한 업데이트 작업을 지속적으로 수행한다. 이를 통해 시스템 이상 탐지, 상태 예측, 효율 최적화 등이 가능하다. 둘째, 테스트 및 시뮬레이션의 경우, 사용자는 CT 모델을 이용하여 실제 PT를 대신해 여러 가상의 상황(예: 과부하, 이상 환경, 충돌 테스트 등)을 재현하고 미래의 반응을 분석할 수 있다. 이러한 시뮬레이션 결과는 물리 시스템의 위험을 줄이고 의사결정의 정확도를 높인다.

DT 시스템에서는 이 두 가지 기능을 지원하기 위해, 다음의 두 종류의 실시간 작업(real-time tasks)이 수행된다. 첫째, 업데이트 작업(update task)은 PT-CT 간 동기화를 유지하기 위한 주기적 데이터 전송 및 반영을 위해 수행된다. 둘째, 추론 작업(inference task)은 사용자의 테스트 요청을 처리하기 위한 실시간 응답형 시뮬레이션을 위해 수행된다. 업데이트 작업과 추론 작업은 서로 밀접하게 관련되어 있다. 예를 들어, 업데이트 작업이 지연되어 CT가 최신 상태로 업데이트되어 있지 않다면, 그 위에서 수행되는 추론 작업의 품질(정확도)이 저하될 수 있다. 따라서 업데이트 작업과 추론 작업의 처리 순서와 주기는 DT 응용의 성능을 결정짓는 핵심 요인이다.

이러한 문제를 다루기 위해 기존 연구에서는 이 업데이트-추론 관계를 명시적으로 고려한 DT 신선도(DT freshness)를 정의하고, 이를 최적화하기 위한 실시간 작업 스케줄링 방안을 제안했다[3]. 기존 방안에서는 추론 작업의 처리율을 저하하지 않는 조건에서 최적의 DT 신선도를 달성하기 위한 작업 스케줄링을 수행한다. 즉, DT 신선도보다 작업 처리율이 더 높은 우선순위를 가지기 때문에,

추론작업에 대한 트래픽 밀도가 증가함에 따라 DT 신선도가 저하하는 한계가 있다. 하지만, DT 신선도가 매우 중요한 응용에서는 추론작업의 처리율의 저하를 감수하더라도 일정 수준 이상의 DT 신선도를 달성할 필요가 있다.

본 논문에서는 가변적인 작업 요청 환경에서 요구되는 신선도를 유지하기 위한 작업 스케줄링 방안을 제안한다. 2장에서는 디지털 트윈 신선도 및 작업 스케줄링과 관련된 기존 연구를 소개한다. 3장에서는 제안 방안이 고려하는 시스템 모델과 요구되는 신선도를 유지하기 위한 작업 스케줄링 방안을 설명한다. 4장에서는 시뮬레이션 결과를 통해 제안 방안의 성능을 검증한다. 마지막으로 5장에서 본 논문의 결론을 맺는다.

## II. Related Works

디지털 트윈은 물리 객체의 동작과 상태를 가상 공간에 실시간으로 반영하기 위한 기술로, 이와 관련된 여러 연구가 진행되었다. 디지털 트윈과 관련하여 기존에 디지털 트윈 모델링 및 구조, 지연 분석, DT 특성 반영 스케줄링 등 다양한 분야에 대한 연구가 이루어졌다. Grieves 등은 디지털 트윈의 기본 구조를 제시하며 물리 객체와 가상 객체 간 데이터 교환 메커니즘을 제안하였다[4]. Tao 등은 모델링, 데이터 수집, 실시간 분석 등 디지털 트윈의 핵심 기능에 대한 체계적인 프레임워크를 제안하였다 [5]. 하지만 이러한 연구는 DT 시스템 내 작업 처리 지연이나 스케줄링 문제를 구체적으로 다루지 않는 한계가 있다. 작업 스케줄링 문제가 디지털 트윈 응용의 성능에 직접적인 영향을 미치기 때문에 이를 다루는 연구도 다수 수행되었다. Kaul 등은 Age of Information (AoI)을 기반으로 갱신 주기와 시스템 성능 간의 관계를 분석하였다 [6]. Sun 등은 무선 네트워크에서 AoI 최적화를 위한 스케줄링 기법을 제안하였다 [7]. 기존 연구들이 다루고 있는 AoI는 디지털 트윈 신선도와 유사한 개념이다. 하지만, AoI는 단일 작업에 대한 처리 지연에 초점을 두고 있는 개념이기 때문에 디지털 환경의 특수한 성질인 갱신 및 추론 작업 간 관계는 고려되지 않는다. 디지털 트윈의 작업 간 관계를 직접 고려한 스케줄링 연구도 일부 제안되었다. Lee 등은 디지털 트윈의 갱신 작업과 추론 작업 간의 제약 관계를 고려한 효율적인 DT 운영 전략을 논의하였으나, 신선도 및 작업 처리율과의 상충관계를 정량적으로 다루지는 않았다 [8]. AoI 기반 기법을 DT 환경에 적용하려는 시도도 있었지만 [9], 대부분 갱신 작업만 고려하고 추론 작업의 처리 지연과 상

호작용은 반영하지 않는다. 즉, 기존 연구는 갱신 및 추론 작업을 포함하는 환경에서 디지털 트윈 신선도와 작업 처리율을 동시에 고려하는 스케줄링 방안이 연구되지 않았다. 본 논문은 이러한 문제를 해결하기 위해, 예상 신선도 기반의 스케줄링 기법을 제안한다.

### III. The Proposed Scheme

이 장에서는 제안방안에 대해 자세히 설명한다. 제안 방안의 목적은 요구되는 디지털 트윈 신선도 유지를 보장하면서 갱신 및 추론 작업을 스케줄링 하는 것이다. 먼저, 제안 방안이 고려하고 있는 시스템 모델과 스케줄링 모델을 설명한다. 다음으로 요구되는 디지털 트윈 신선도를 유지할 수 있는 구체적인 방안을 서술한다.

제안방안이 고려하고 있는 디지털 트윈 시스템의 구조는 그림 1과 같다.  $K$ 개의 물리 객체(Physical Object; PO)가 현장에서 운용되고 있으며, 서버는 각 물리 객체에 대응하는  $K$ 개의 디지털 객체(Digital Object; DO)를 관리하고 있다. 각 물리 객체에서 수집된 데이터는 주기적으로 서버에 보고되며, 서버는 이 정보를 통해 대응하는 디지털 객체의 상태를 최신화하는 갱신 작업(update task)을 수행한다. 각각의 물리 객체는 서로 다른 응용을 위해 운용되고 있으며, 이에 따라 각각의 디지털 객체는 서로 다른 요구 신선도를 가진다. 물리 객체와 서버로 구성되는 디지털 트윈 시스템의 외부에는 각 응용과 관련된 원격 사용자들이 추론 작업 (inference task) 요청을 수행한다. 서버는 사용자의 추론 작업 요청에 따라 디지털 객체로부터 상태 조회, 상태 예측, 시뮬레이션 결과 등을 응답한다.

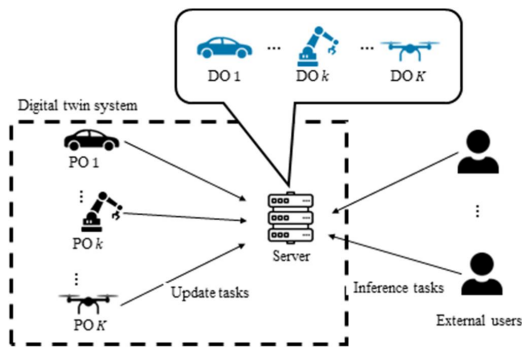


Fig. 1. Digital Twin System

디지털 객체  $k$ 에 대해 수행되는 컴퓨팅 작업은 다음과 같이 모델링 될 수 있다.

$$\tau_{k,i} = (t_{k,i}^{arr}, T_{k,i}^{comp}, T_{k,i}^{dead}) \quad (1)$$

(1)에서  $i$ 는 작업의 순서를 의미한다. 디지털 객체  $k$ 에 대한 작업 중 갱신 작업이 가장 먼저 수행된다고 가정하면  $\tau_{k,0}$ 는 디지털 객체  $k$ 에 대한 갱신 작업을 의미한다.  $t_{k,i}^{arr}$ 과  $T_{k,i}^{comp}$ 는 각각 작업의 도착 시간과 작업이 수행되는데 필요한 컴퓨팅 지연을 의미한다. 도착 시간은 갱신 작업의 경우 물리 객체의 데이터가 서버에 도착한 시점, 추론 작업의 경우 사용자 요청이 서버에 도착한 시점을 의미한다. 컴퓨팅 지연은 작업의 크기, 디지털 객체의 복잡도, 그리고 서버의 성능에 좌우된다. 즉, 작업의 크기가 크거나 디지털 객체의 복잡도가 클수록 컴퓨팅 지연이 커지고 서버의 성능이 높을수록 컴퓨팅 지연이 작아진다.  $T_{k,i}^{dead}$ 는 작업의 상대적인 제한 시간을 의미한다. (1)로부터 작업의 절대적인 제한 시간  $t_{k,\xi}^{dead} = t_{k,\xi}^{arr} + T_{k,\xi}^{dead}$ 를 도출할 수 있다.

갱신 작업과 추론 작업은 발생 주기, 컴퓨팅 지연, 그리고 제한 시간 면에서 차이점을 가진다. 갱신 작업의 경우 물리 객체에서 일정 시간 동안 축적된 데이터에 의해 주기적으로 수행된다. 따라서 축적된 데이터를 디지털 객체에 반영하기 위한 상대적으로 긴 컴퓨팅 지연을 가지며 갱신 작업의 절대적인 제한 시간은 다음 갱신 작업이 생성되는 시점이다. 즉, 갱신 작업의 상대적인 제한 시간은 갱신 작업이 생성되는 주기와 같다. 반면, 추론 작업의 경우, 외부 사용자에게 의해 임의의 시간에 요청되며, 사용자 요청에 대한 실시간 응답이 요구된다. 따라서 디지털 객체의 갱신 주기 동안 다수의 추론 작업이 생성될 수 있으며, 갱신 작업보다 작은 상대적인 제한 시간을 가진다.

그림 2는 서버에서 디지털 객체  $k$ 에 대한 일련의 작업이 수행되는 과정을 시계열로 가시화한 예시이다. 서버는 다수의 디지털 객체에 대한 갱신 및 추론 작업을 수행하기 때문에, 어떤 작업  $\tau_{k,i}$ 는 큐 지연으로 인해 도착 시간에 즉시 수행되지 못할 수 있다. 그림 2는 디지털 객체  $k$ 에 대한 갱신 작업의 도착 시간부터 제한 시간까지의 작업 처리를 보여주며,  $t_{k,i}^{start}$ 와  $t_{k,i}^{end}$ 는 각각  $\tau_{k,i}$ 의 수행이 서버에서 시작된 시점과 종료된 시점을 의미한다. 그림에서, 첫 번째 추론 작업인  $\tau_{k,1}$ 는 작업 도착 즉시 수행되었고, 두 번째 추론 작업인  $\tau_{k,2}$ 는 도착 후 다른 디지털 객체를 위한 갱신 및 추론 작업으로 인한 큐 지연을 대기한 후 수행되었다. 그림처럼, 갱신 주기 동안 한 디지털 객체에 대한 작업은 하나의 갱신 작업과 여러 개의 추론 작업이 수행될 수 있다.

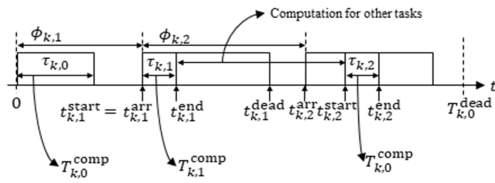


Fig. 2. Update and Inference Tasks

신선도는 디지털 객체에 대한 갱신 작업과 추론 작업 사이의 관계를 반영하여 정의되는 디지털 트윈 응용의 고유한 성능지표이다. 디지털 객체의 상태 변화를 그림 3과 같이 구분할 수 있다. 특정 시점에 도착한 갱신 작업의 제한 시간이 다음 갱신 작업이 도착하는 시점임을 고려했을 때, 갱신 작업의 주기를  $T_{k,0}^{dead}$ 로 정의할 수 있다. 자연수  $x$ 에 대해,  $xT_{k,0}^{dead}$ 에 도착한 갱신 작업을  $\tau_{k,0}(x)$ 라고 하면,  $\tau_{k,0}(x)$ 는 시간 간격  $((x-1)T_{k,0}^{dead}, xT_{k,0}^{dead})$  동안 물리 객체에서 수집된 데이터를 포함한다.  $\tau_{k,0}(x)$ 에 의해 갱신된 디지털 객체  $k$ 의 상태를  $S_k(x)$ 라고 하면,  $\tau_{k,0}(x)$ 의 처리의 완료 여부에 따라 디지털 객체  $k$ 의 상태를 구분할 수 있다. 즉,  $\tau_{k,0}(x)$ 가 도착한 시점에 디지털 객체  $k$ 는  $\tau_{k,0}(x-1)$ 에 의해 갱신된 상태이기 때문에 디지털 객체의 상태는  $S_k(x-1)$ 이고,  $\tau_{k,0}(x)$ 가 처리된 이후에 디지털 객체의 상태는  $S_k(x)$ 로 갱신된다. 그림 3의 상단은  $\tau_{k,0}(x)$ 가 도착한 시점에 처리되는 경우를 보여준다. 이 경우, 시점  $t_1$ 에 디지털 객체의 상태가  $S_k(x-1)$ 에서  $S_k(x)$ 로 갱신되기 때문에 이후에 도착하는 추론 작업은  $S_k(x)$ 에 의해 처리된다. 반면에 그림 3의 하단은  $\tau_{k,0}(x)$ 가 다른 작업의 처리를 위해 지연된 경우를 보여준다. 이 경우,  $\tau_{k,0}(x)$ 의 처리가 완료되는 시점  $t_2$ 에 디지털 객체의 상태가  $S_k(x-1)$ 에서  $S_k(x)$ 로 갱신된다. 그 결과 시점  $t_2$  이전에 처리되는 추론 작업은 이전 상태인  $S_k(x-1)$ 에 의해 처리되고, 시점  $t_2$  이후에 처리되는 추론 작업은 최신 상태인  $S_k(x)$ 에 의해 처리된다. 추론 작업은 디지털 객체의 실시간 상태에 의해 처리되는 것을 기대하며 생성된다는 점을 고려했을 때, 사용자들이 받게 되는 추론 결과의 품질은 그림 3 상단의 경우가 그림 3 하단의 경우보다 더 높다. 이러한 관점에서 특정 갱신 주기에서의 디지털 트윈 신선도는 전체 추론 작업 중 갱신 작업 이후에 처리되는 추론 작업의 비율로 정의될 수 있다. 구체적으로, 본 논문에서는 주어진 갱신 간격  $((x-1)T_{k,0}^{dead}, xT_{k,0}^{dead})$ 에 대해 디지털 트윈 신선도  $F_k(x)$ 를 다음과 같이 정의한다.

$$F_k(x) = \frac{n(x)}{n_t(x)} \quad (2)$$

(2)에서  $n(x)$ 는 갱신 간격  $((x-1)T_{k,0}^{dead}, xT_{k,0}^{dead})$  동안 처리된 전체 추론 작업의 수 이고,  $n_t(x)$ 는 해당 디지털 객체에 대한 갱신 작업이 완료된 시점 이후에 처리된 추론 작업의 수를 의미한다. 즉, 전체 추론 작업 중 최신 상태로 갱신된 디지털 객체를 통해 처리된 추론 작업의 수를 디지털 트윈 신선도로 정의한다.

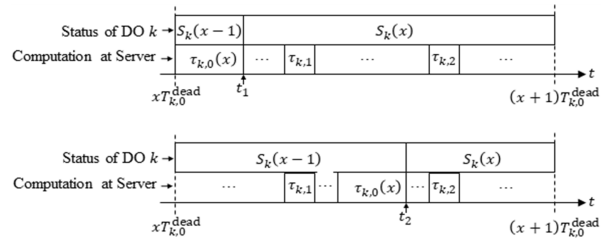


Fig. 3. Concept of Digital Twin Freshness

서버에서의 실시간 작업 스케줄링 모델은 갱신 작업과 추론 작업을 위한 두 개의 작업 큐를 통해 이루어진다. 그림 4는 제안 방안이 고려하고 있는 실시간 작업 스케줄링 모델을 보여준다. 서버로 요청된 작업은 작업 분류기 (Task classifier), 동종 작업 간 스케줄러 (Intra-task scheduler), 그리고 이종 작업 간 스케줄러 (inter-task scheduler)를 통해 프로세서에서 처리된다. 첫 번째 단계인 작업 분류기에서는 요청된 작업이 갱신 작업인지 추론 작업인지 분류하여 해당하는 작업 큐에 삽입한다. 이 과정은 응용 헤더의 플래그 등을 이용하여 간단하게 수행될 수 있다. 두 번째 단계인 동종 작업 간 스케줄러에서는 각 작업 큐에 삽입된 동종 작업 간 순서를 결정한다. 이 단계에서는 두 가지 요소가 고려된다. 첫째는 디지털 객체 간 우선순위이다. 모든 디지털 객체가 동일한 응용에 포함되는 경우 응용 관리자에 의해 각 디지털 객체의 우선순위가 결정될 수 있다. 각 디지털 객체가 서로 다른 응용에 포함되는 경우, 응용 간 우선순위가 디지털 객체 간 우선순위에 그대로 반영될 수 있다. 둘째는 작업의 제한 시간이다. 즉, 제한 시간이 가장 적게 남은 작업을 가장 먼저 처리해야 한다. 서버에서 제한 시간을 초과한 작업은 처리가 거부되기 때문에, 작업 처리율을 높이기 위해 반드시 고려되어야 하는 요소이다. 한 가지 유의해야 할 점은 잔여 제한 시간의 양은 작업의 도착 시각과 무관하다는 점이다. 즉, 제한 시간이 짧은 작업은, 먼저 도착했지만 제한 시간이 훨씬 긴 다른 작업보다 먼저 처리될 수 있다. 이는 각 디지털 객체에 대한 갱신 작업의 주기나 추론 작업에 요구되는 응답

시간이 서로 다를 수 있음을 의미한다. 본 논문에서는 모든 디지털 객체의 우선순위가 모두 동일하다고 가정하고, 잔여 제한 시간만을 고려하여 동종 작업 간 스케줄링을 결정한다. 작업 스케줄링의 마지막 단계는 이종 작업 간 스케줄링 단계이다. 기존 방안에서는 이 단계에서 작업 처리율을 저하하지 않는 조건으로 스케줄링을 수행한다. 따라서 작업 요청 빈도가 증가하면 디지털 트윈 신선도  $F_k(x)$ 가 저하할 우려가 있다. 반면, 제안 방안은 이 단계에서 예상되는 디지털 트윈 신선도를 점검하고, 요구되는 값보다 저하되지 않도록 스케줄링을 수행한다.

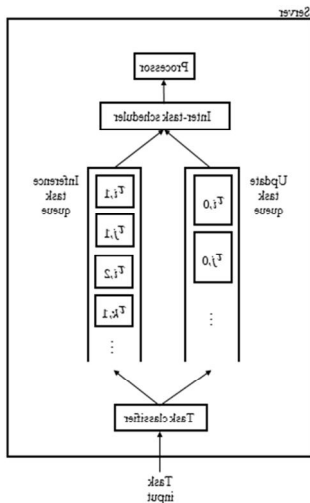


Fig. 4. Task Scheduling Architecture

제안 방안이 디지털 트윈 신선도를 점검하는 구체적인 방법은 다음과 같다. (2)에서 볼 수 있는 것처럼, 디지털 트윈 신선도는 한 주기 동안 수행된 전체 추론 작업에 대한 갱신 작업이 처리된 후 수행된 추론 작업의 비율을 의미한다. 예를 들어, 한 주기 동안 10개의 추론 작업이 수행되었고, 4번째와 5번째 추론 작업 사이에 갱신 작업이 수행되었다면, 10개 중 6개의 추론 작업이 갱신된 최신 CT에 의해 처리되었기 때문에  $F_k(x) = 0.6$  이다. 즉, 갱신 작업의 수행 시점에 의해서 달성되는 디지털 트윈 신선도가 결정된다. 하지만, 디지털 트윈 신선도  $F_k(x)$ 는 한 주기 동안 모든 작업이 수행된 이후에 정의될 수 있기 때문에, 이종 작업 간 스케줄링 단계에서 이를 정확히 측정할 수 없다. 따라서 제안 방안은 추론 작업의 도착 모델을 기반으로 예상되는 디지털 트윈 신선도를 측정하고, 이를 통해 갱신 작업의 수행 시점을 결정한다.  $x$  번째 주기의 시간 간격을  $((x-1)T_{k,0}^{dead}, xT_{k,0}^{dead})$ , 현재 시점을  $t$  라고 할 때,  $I(t)$ 를 이미 수행된 추론 작업의 수,  $E[U(t)]$  앞으로 수행될 것으로 예상되는 추론 작업의 수라고 하자.

시점  $t$ 에 스케줄링되는 작업의 종류에 따라 두 가지 경우를 고려할 수 있다. 첫째, 갱신 작업을 스케줄링 하는 경우, 달성될 것으로 예상되는 디지털 트윈 신선도는  $E[U(t)]/(I(t)+E[U(t)])$ 이다. 이 경우는 현재 달성 가능한 신선도가 요구되는 신선도보다 큰 값인지는 판단할 수 있지만, 시점  $t$ 에서의 갱신 작업 지연이 요구되는 신선도 달성 실패를 야기하는지 판단할 수 없다. 둘째, 추론 작업을 스케줄링 하는 경우, 달성 가능한 최대 신선도는 다음과 같다.

$$(E[U(t)] - 1)/(I(t) + E[U(t)]) \quad (3)$$

즉, 시점  $t$ 에 스케줄링되는 추론 작업이 끝난 직후 갱신 작업을 처리함으로써 최대 신선도를 달성할 수 있다. 이 때, (3)이 요구되는 신선도보다 작은 값인 경우, 현재 갱신 작업을 수행해야 한다는 신호로 해석할 수 있다. 따라서 제안 방안은 (3)이 요구되는 신선도보다 작아지는 경우 갱신 작업을 처리함으로써 요구되는 신선도 달성을 보장할 수 있다.

예상 디지털 트윈 신선도를 정확히 계산하기 위해서는 추론 작업의 요청 시점이 미리 알려져 있어야 한다. 예를 들어, 추론 작업이 주기적으로 요청되는 경우,  $E[U(t)]$ 가 정확히 계산될 수 있다. 하지만 일반적으로 추론 작업은 사용자에게 의해 임의로 요청된다. 이 경우, 추론 작업 요청 모델에 기반하여 확률적으로 디지털 트윈 신선도 보장률을 계산할 수 있다. 예를 들어, 추론 작업의 요청이 도착률  $\lambda$ 의 포아송 프로세스를 따르고, 요구되는 디지털 트윈 신선도 보장률이  $p$  ( $0 < p \leq 1$ ) 라고 가정하자. 이 경우, 포아송 분포의 누적 분포 함수를 이용하여  $p$  이상의 확률로 요청될 것으로 예상되는 최대 추론 작업 수를 구할 수 있다. 즉, 다음을 만족하는  $m$ 을 구할 수 있다.

$$P(E[U(t)] \leq m) = \sum_{i=0}^m \frac{\lambda (T_{k,0}^{dead} - t)^i e^{-\lambda(T_{k,0}^{dead} - t)}}{i!} \geq p \quad (4)$$

즉, (3)의  $E[U(t)]$  대신에 (4)를 만족하는  $m$ 을 대입하면  $p$ 이상의 확률로 디지털 트윈 신선도를 보장할 수 있다.

#### IV. Performance Evaluation

본 장에서는 제안 방안의 성능 평가를 위해 수행한 시뮬레이션 결과를 분석한다. 시뮬레이션은 작업 처리율에 우선 순위를 둔 기존 방안[3]과 함께 수행되었으며, 비교를 위해 갱신 작업 우선 (update-first) 방안과 추론 작업 우선 (inference-first) 방안을 포함하였다. 갱신 작업 우선 방안에서는 모든 갱신 작업을 처리한 후 추론 작업이 수행된다. 따라서 갱신 작업이 수행되는 동안 요청되는 추론

작업은 모두 거부된다. 반면에, 추론 작업 우선 방안은 추론 작업 큐가 비어있는 경우에만 갱신 작업을 수행한다. 시뮬레이션은 다음과 같은 환경에서 수행되었다. 작업의 대상이 되는 가상 객체는 10개이다. 갱신 작업과 추론 작업 처리 지연은 각각 2ms와 0.5ms이며, 갱신 작업 주기는 100ms이다. 요구되는 신선도는 0.7로 설정하였으며, 전체 가상 객체에 대한 통합 작업 도착율은 1task/ms 이다. 표 1은 시뮬레이션 환경을 보여준다.

Table 1. Simulation setting

Parameter	Value
Number of CTs	10
Update task delay	2 ms
Inference task delay	0.5 ms
Update task interval	100 ms
Desired DT freshness	0.7
Task arrival rate	1 task/ms

그림 5는 작업 도착율 증가에 따른 디지털 트윈 신선도를 보여준다. 갱신 작업 우선 방안은 모든 갱신 작업을 수행한 후에 추론 작업 요청을 받아들이기 때문에, 가장 높은 디지털 트윈 신선도를 달성한다. 반면, 추론 작업 우선 방안은 작업 도착율이 증가할수록 갱신 작업의 처리가 지연되므로 디지털 트윈 신선도가 감소한다. 기존 방안의 경우, 작업 도착율이 낮을 때는 갱신 작업 우선 방안보다 높은 디지털 트윈 신선도를 달성하지만, 추론 작업 우선 방안과 마찬가지로 작업 도착율이 증가할수록 디지털 트윈 신선도가 감소한다. 이는 작업 처리율에 우선 순위가 있기 때문에, 작업 도착율이 증가할수록 추론 작업의 처리가 우선되기 때문이다. 제안 방안도 기존 방안과 마찬가지로 작업 도착율이 증가할수록 디지털 트윈 신선도가 감소하지만, 요구되는 신선도 수준을 유지한다. 이는 (3)을 기반으로 디지털 트윈 신선도 저하가 예상되는 경우 갱신 작업을 수행하기 때문이다.

그림 6은 작업 도착율 증가에 따른 작업 거부율을 보여준다. 갱신 작업 우선 방안의 경우, 갱신 작업을 수행하는 동안 요청되는 모든 추론 작업이 거부되기 때문에 일정한 비율의 작업이 반드시 거부된다. 반면에 추론 작업 우선 방안은 모든 작업을 처리하지 못할 정도로 작업 도착율이 높지 않은 경우 거의 모든 작업이 성공적으로 처리된다. 기존 방안의 경우, 추론 작업이 거부되지 않는 조건에서 갱신 작업의 처리 시점을 조정함으로써 디지털 트윈 신선도를 최적화하기 때문에, 추론 작업 우선 방안과 거의 동일한 작업 거부율을 보인다. 제안 방안의 경우, 작업 도착율이 낮은 경우, 작업 거부가 거의 발생하지 않지만, 작업 도

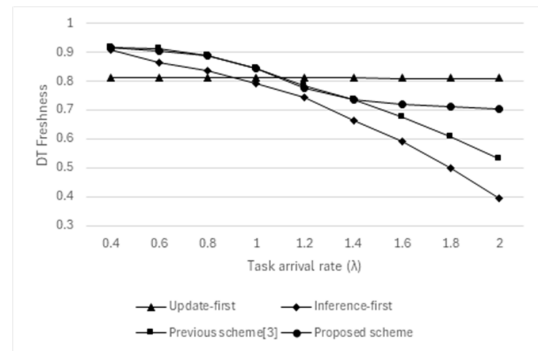


Fig. 5. DT freshness according to task arrival rate

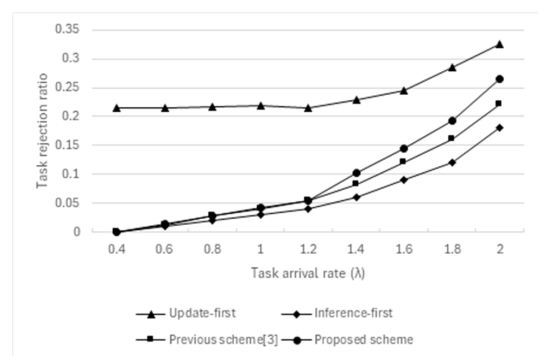


Fig. 6. Task rejection ratio according to task arrival rate

작율이 증가하면서 작업의 거부가 발생한다. 이는 (3)이 요구되는 신선도보다 작아지는 경우 수행되는 갱신 작업의 스케줄링으로 인해 해당 시점에 추론 작업 큐에 존재하는 작업의 거부가 발생할 수 있기 때문이다.

시뮬레이션 결과에서 볼 수 있는 것처럼, 기존 방안은 작업 처리율에 우선순위를 두고 스케줄링을 수행하기 때문에 높은 작업 도착율 환경에서 디지털 트윈 신선도를 희생함으로써 작업 거부율을 최소화한다. 이는 작업 처리율이 중요한 일반적인 응용에 적합할 수 있지만, 디지털 트윈 응용이 일정 수준 이상의 디지털 트윈 신선도를 요구하는 경우에 대응할 수 없다. 반면에, 제안 방안은 작업 스케줄링 과정에서 실시간으로 달성 가능한 신선도를 계산하고, 이를 바탕으로 갱신 작업 수행 시점을 가변적으로 조정함으로써 요구되는 디지털 트윈 신선도를 보장한다. 이 과정에서 기존 방안보다 높은 작업 거부율을 보이지만, 요구되는 신선도의 유지가 핵심 지표인 디지털 트윈 응용에서는 제안 방안이 기존 방안보다 더 적합한 작업 스케줄링 방식이다.

### V. Conclusions

본 논문에서는 디지털 트윈 시스템에서 디지털 트윈 신선도를 보장하기 위해, 예상 신선도 기반의 작업 스케줄링

방안을 제안하였다. 기존 방안은 작업 처리율에 우선순위를 두기 때문에, 작업 도착율이 낮은 경우 높은 디지털 트윈 신선도를 달성할 수 있지만, 작업 도착율이 높은 경우, 추론 작업의 처리를 위해 갱신 작업을 지연시킴으로써 디지털 트윈 신선도의 저하를 야기한다. 이를 해결하기 위해 제안 방안은 예상되는 추론 작업 수를 기반으로 달성 가능한 디지털 트윈 신선도를 계산하고, 이를 통해 갱신 작업 수행 시점을 결정한다. 즉, 디지털 트윈 신선도가 요구되는 값을 달성하지 못할 것으로 판단되는 경우 갱신 작업을 우선 처리함으로써 디지털 트윈 신선도 저하를 예방한다. 시뮬레이션 결과는 제안 방안이 제공하지 못하는 디지털 트윈 신선도 보장 기능을 제공함을 보여준다. 다만, 제안 방안은 신선도 보장을 위해 일부 작업의 거부를 필요로 한다. 따라서 제안 방안은 작업 처리율 보다는 디지털 트윈 신선도 보장이 더 중요한 요소로 작용하는 응용에 적합하다. 향후 연구에서는 일정 수준 이상의 디지털 트윈 신선도가 요구되는 환경에서 작업 거부율을 줄일 수 있도록 작업 요청 빈도와 작업 제한 시간을 조정할 수 있는 프레임워크에 대한 연구가 필요하다. 또한, 응용 간 또는 디지털 객체 간 상이한 중요도를 가지는 디지털 트윈 시스템에서 신선도 보장을 위한 작업 스케줄링 방안을 연구할 계획이다.

## ACKNOWLEDGEMENT

This was supported by Korea National University of Transportation in 2024.

## REFERENCES

- [1] O. Hashash, C. Chaccour, W. Saad, T. Yu, K. Sakaguchi and M. Debbah, "The Seven Worlds and Experiences of the Wireless Metaverse: Challenges and Opportunities," in *IEEE Communications Magazine*, vol. 63, no. 2, pp. 120-127, February 2025.
- [2] W. Saad et al., "Artificial General Intelligence (AGI)-Native Wireless Systems: A Journey Beyond 6G," *Proceedings of the IEEE*, early access.
- [3] C. Kim, M. Chehimi, M. Jung and W. Saad, "Real-Time Task Scheduling for Digital Twin Edge Network," 2023 IEEE Global Communications Conference, Kuala Lumpur, Malaysia, 2023.
- [4] M. Grieves and J. Vickers, "Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems," Springer, 2017.
- [5] F. Tao, H. Zhang, A. Liu and A. Y. C. Nee, "Digital Twin in Industry: State-of-the-Art," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2405-2415, April 2019.
- [6] S. Kaul, R. Yates and M. Gruteser, Real-time status: How often should one update, 2012 Proceedings IEEE INFOCOM, Orlando, FL, USA, 2012.
- [7] Y. Sun, E. Uysal, R. Yates, C. Singh, and S. Kaul, "Age of Information: A New Metric for Information Freshness," *Synthesis Lectures on Communication Networks*, Morgan & Claypool, 2019.
- [8] S. Vestal, "Preemptive Scheduling of Multi-criticality Systems with Varying Degrees of Execution Time Assurance," 28th IEEE International Real-Time Systems Symposium (RTSS 2007), Tucson, AZ, USA, 2007.
- [9] J. Lee, H. Davari, J. Singh, and V. Pandhare, "Industrial Artificial Intelligence for Industry 4.0," *Manufacturing Letters*, vol. 18, pp. 20-23, 2018.

## Authors



Taehun Yang received his B.S. degree in Computer Engineering from Chungnam National University in 2014 and obtained his Ph.D. degree in Computer Engineering from the same university in 2022.

Taehun Yang worked as a postdoctoral researcher at Kongju National University from 2022 to 2023. In 2023, he moved to Gyeongbuk National University (formerly Andong National University) and is currently serving as an assistant professor in the Computer Science and Engineering Major, School of IT Convergence Engineering. His research topics mainly include computer communication, mobile computing, wireless communication, the Internet of Things, and digital twins.



Cheonyong Kim received the B.S., M.S., and Ph.D. degrees in computer science and engineering from Chungnam National University, Daejeon, South Korea, in 2013, 2015, and 2019, respectively.

Cheonyong Kim is currently an Assistant Professor at the Department of Computer Software at Korea National University of Transportation, South Korea. From March 2022 to August 2024, he was a Postdoctoral Researcher with the Department of Electronics and Information Engineering, Sejong University, South Korea. He was also a Visiting Researcher with the Bradley Department of Electrical and Computer Engineering, Virginia Tech, VA, USA. From July 2019 to September 2021, he was a Postdoctoral Researcher with the Korea Institute of Science and Technology Information (KISTI), Daejeon, South Korea. His research interests include wireless communication systems, mobile edge computing (MEC), and digital twins.