



(T1)를 시계열 DB와 IoT 관리 서버에 전송한다. 데이터 결측 확인 SW는 일정 주기(T2)마다 IoT 관리 서버에 디바이스의 주기 데이터를 요청하고, 시계열 DB에 T2 기간 동안의 센서 데이터 전송 시각 리스트를 요청한다. 요청한 데이터를 바탕으로 결측 판단에 필요한 데이터를 계산하여 다시 시계열 DB에 저장한다.

### 3. 구현

#### 3.1 IoT 디바이스 데이터 결측 탐지 알고리즘

결측 탐지 알고리즘은 일정 기간의 데이터 전송 주기의 평균과 해당 기간의 결측 판단 결과를 계산한다. 이를 위해서는 다음의 2가지 파라미터(결측 탐지 주기, 결측 판별 기준)를 필요로 한다. T2를 결측 탐지 주기로, K를 결측 판별 기준이라고 정의한다.

알고리즘은 T2주기마다 각 센서에 대해 다음의 작업을 실행한다. 먼저, IoT 관리 서버에 HTTP GET request를 통해 사용자가 지정한 데이터 전송주기(T1)을 조회한다. InfluxDB에는 ‘마지막으로 센서 데이터가 전송된 시각(TL)’으로부터 현재 시각까지의 데이터를 조회하고 이 중 time 필드만 추출하여 리스트에 저장한다. T2 주기 동안의 타임스탬프 리스트와, 사용자가 지정한 데이터 전송주기(T1), 결측 판별 기준 K를 알고 있으므로 이를 비교하여 T2 기간 동안의 데이터 결측 발생 여부를 판단할 수 있다.

T2 주기 동안의 타임스탬프 리스트와 마지막으로 센서 데이터가 전송된 시각을 알고있으므로, T2 주기 동안 센서 데이터가 전송된 시간의 간격 정보를 얻을 수 있다. 데이터 전송 시간 차를 저장한 배열의 원소들을 모두 합하여 평균을 내면, T2 주기 동안의 데이터 전송 주기의 평균값을 얻을 수 있다. 데이터 전송 주기의 평균값과 T1과의 차이를 구하여 백분율로 변환한 후, 그 차이가 K보다 작거나 같으면 결측이 발생하지 않은 것으로 판단, K보다 크면 결측이 발생했다고 판단한다.

그림 2는 2개의 온도 센서 IoT 디바이스로 데이터 결측 탐지 알고리즘을 실험한 결과를 Grafana를 통해 시각화 한 것이다. 그래프의 가로축은 시간, 세로축은 데이터를 의미한다. 위 그래프는 데이터 결측 판단 결과를 나타낸 그래프로, 결측이 발생하면 1, 결측이 발생하지 않았으면 0의 값을 표시한다. 아래 그래프는 데이터 전송 주기의 평균값을 나타낸 그래프이다.

실험 조건은 T2: 300초, K: 10, 2개의 디바이스 모두 T1: 60초이다. 알고리즘 테스트를 위해 Env08(노란 디바이스)의 전원을 일정 기간 차단 후 공급하였고, Env02(초록 디바이스)는 정상 동작을 유지했다. 결과적으로, Env02 디바이스는 전원이 차단된 기간 동안 데이터 전송을 하지 못했으므로 데이터 전송주기의 평균값이 증가하고, 결측 판단 결과는 1이 나오는 것을 확인하였다. 반대로 Env08 디바이스는 정상 동작을 유지하였으므로 데이터 전송주기의 평균값이 T1과 비슷하게 일정하게 계산되고, 결측 판단 결과는 0이 나오는 것을 확인하였다.

### 4. 결론 및 향후 연구 방향

본 논문에서는 실시간으로 IoT 디바이스의 데이터 결측을 탐지하는 알고리즘과 시스템을 제안하였다. 소개한 방법은 서버에 부담을 최소화하는 방법으로, IoT 데이터 결측을 준 실시간으로 검사할 수 있다는 장점이 있다. 추후 데이터 결측 탐지 알고리즘 실행 도중, 디바이스의 데이터 전송주기가 변경되는 경우를 고려하여 알고리즘을 고도화하는 추가 연구가 필요할 것으로 생각된다.

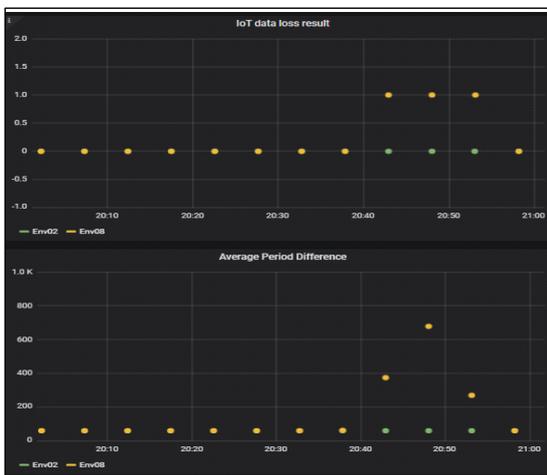
### Acknowledgement

이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2021-0-00471, 모델링 & 최적화 기반 오류-free 정보인프라 자율제어 기술 개발)

### 참고문헌

[1] B. Fekade, T. Maksymyuk, M. Kyryk and M. Jo, "Probabilistic Recovery of Incomplete Sensed Data in IoT," in IEEE Internet of Things Journal, vol. 5, no. 4, pp. 2282-2292, Aug. 2018, doi: 10.1109/JIOT.2017.2730360.

[2] Gangadhar S, Sowmya. (2020). The Real Time Environmental Time Series Data Analysis Using Influx DB. International Journal of Advanced Scientific Innovation, 1(1). <https://doi.org/10.5281/zenodo.4641703>



[그림 2] IoT 디바이스 데이터 결측 탐지 알고리즘 실행 결과