

수도사업장 데이터 결측 해결 방안에 관한 연구

이동훈, 박시현, 김종립, 김대욱, 신강욱, 홍성택

한국수자원공사 연구원

{dhhero, seanpark, kjr8963, therealone, gwshin, sthong}@kwater.or.kr

A Study on the Prevention of Missing Data in Water Treatment Plant

Lee Dong Hoon, Park Si Hyun, Kim Jong Rib, Kim Dae Wook, Shin Gang Wook, Hong Sung Taek

Korea Water Resources Corporation

요 약

본 논문은 4차 산업혁명 시대의 도래로 중요성이 증대되고 있는 데이터의 질적 확보를 위해 수도사업장 전송 데이터 결측 문제에 대한 해결 방안을 제시하였다. 수도사업장 데이터 전송체계 현황 분석 및 문제점 분석을 통해 수도사업장 데이터 전송체계 개선방안을 제시하였다. 개선방안은 전송체계를 간소화하는 방안, 전용 DB를 도입하는 방안, 데이터전송 및 결측 복구기능을 분리하는 3가지 방안이며, 제시된 방안들에 대한 장단점 분석을 통해 최적 방안을 도출하는 연구를 수행하였다.

I. 서 론

4차 산업혁명 시대 데이터는 21세기 쌀과 원유로 비유될 정도로 중요성이 높아지고 있다. 정보기술(IT)은 데이터 활용, 데이터 기반 기술로 비즈니스 축이 이동하고 있으며 빅데이터, AI, 사물인터넷 등 신기술의 등장은 데이터에 대한 관심을 더욱 높이고 있다. 데이터는 그 자체보다 데이터를 활용한 가치 창출이 중요하며 데이터 활용을 통한 고급분석, 미래예측 능력은 기업과 국가의 경쟁력에도 직결되는 실정이다. 기업뿐만 아니라 국가에서도 데이터 기반의 의사결정 체계를 기본 인프라로 구축하려는 움직임이 활발함에 따라 데이터의 양적 확보 외에 양질의 데이터를 위한 품질관리에도 관심이 집중되고 있다. 그러나 데이터 품질관리를 위한 노력과 비용투자는 데이터 활용에 비해 매우 미흡한 실정이다.[1][2]

수도사업장에서는 대부분 원격감시제어를 통해 깨끗한 수돗물을 생산하고 있다. 원격감시제어는 현장 설비와 센서 등의 데이터를 감시하고, 중앙 조정실로 전송되는 데이터를 기반으로 설비 제어를 원격에서 운영한다는 의미이다. 현장이 아닌 원격에서 시설 감시제어가 이뤄지므로 오결측 없는 데이터는 안정적인 사업장 운영에 필수적이다. 수도사업장에 활용되는 데이터는 현장에서 생성되어 다양한 경로를 통해 최종적으로는 데이터 베이스(DB)에 저장된다. 빅데이터 기술의 발전으로 최근 데이터베이스에 축적된 데이터를 활용하여 수도사업장에 AI를 활용한 자율운영체계를 구축하려는 연구도 활발해지고 있다. 그러므로, AI, 빅데이터 기반의 의사결정시스템을 도입하기 위해서 결측 없는 데이터 확보를 위한 체계적인 전송체계가 필요하다.[3]

수도사업장의 데이터 생성-전송-저장으로 이어지는 일련의 과정은 다양한 시스템과 프로그램을 거치는 복잡한 구조이므로 데이터 결측 발생 가능성이 높다. 데이터 전송체계를 단순화하고 결측시 데이터를 복구할 수 있는 방안 등 다각적인 연구가 필요한 시점이다.

따라서, 본 연구에서는 현재 수도사업장 데이터 전송체계 현황을 분석하고 문제점을 도출하여 데이터 결측 문제를 해결할 수 있는 개선방안을 제시하고자 한다.

II. 본론

2.1 수도사업장 데이터 전송체계 현황 및 문제점 분석

광역 수도사업장에서는 현장 시설물, 계측기기의 상태를 원격에서 감시 제어하고 있다. 각 시설물, 계측기기의 데이터는 현장설비에서 생성되어 그림1과 같은 전송체계를 통해 본사DB로 전송되고 있다.

일반적으로 각 설비의 상태는 현장의 감시제어설비(PLC, DCS 등)를 거쳐 중앙조정실의 SCADA서버로 전송되고 운영근무자는 HMI를 통해 사업장을 운영할 수 있다. HMI에 표시되는 데이터는 실시간DB(RTDB)에 매분마다 저장된다. 관리단 RTDB에 저장된 데이터는 본부 RTDB로 전송된 후 DB 전송프로그램을 통해 본부DB에 저장되며, 본부DB에 저장된 데이터는 최종적으로 본사DB로 전송 및 저장되는 전송체계이다.

이와 같이 하나의 데이터가 현장에서 생성되어 본사DB에 저장되기까지 다양한 시스템과 프로그램 등 복잡한 전송과정을 거치게 되며, 이러한 전송과정에서 여러 원인에 의해 데이터의 결측이 발생한다. 데이터의 결측은 크게 ①시스템 이중화 문제, ②전송프로그램 문제, ③시스템간 시간 동기화 문제, ④결측 데이터 복구문제로 나눌 수 있다.

각각의 원인은 다음과 같다. 첫째, 시스템 이중화 문제는 SCADA서버가 이중화로 운영되는 환경에서, 유지보수 등의 SCADA서버 작업시 정상운영중인 서버로 데이터전송 기능을 절체하지 않아 발생한다. 둘째, 전송프로그램의 오류로 지사-본부 RTDB간 데이터전송 서비스가 원활하지 않는 경우가 있다. 셋째, SCADA서버 및 DB서버간의 시간 동기화가 되지 않아 데이터 결측이 발생한다. 넷째, 장기간 결측시 시스템 부하 등에 의한 데이터복구가 어려운 문제가 있다.

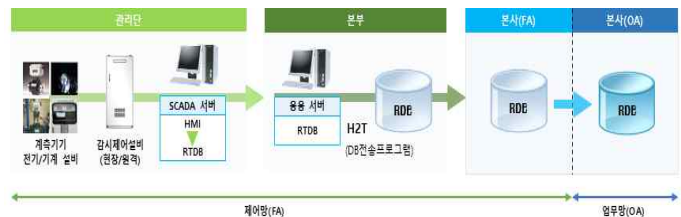


그림 1 수도사업장 데이터 전송체계 현황

2.2 데이터 결측 해결 방안

본 연구에서는 데이터 결측문제를 해결할 수 있는 3가지 방안에 대해 제시하고자 한다.

가. 1안 : 전송체계 간소화

데이터 전송상 오결측 장애는 고장점이 많아지면 발생할 확률이 높으므로 장애 요인을 줄이는 것이 필요하다. 그림1과 같이 관리단에서 본부로의 데이터 전송은 RTDB간 데이터 전송으로 이뤄진다. RTDB간 통신을 위해서는 백그라운드 프로그램이 동작하는 상황에서 RTDB간 1분데이터를 전송하여 많은 네트워크 트래픽을 발생시킨다. 이와 같은 상황에서 백그라운드 프로그램의 서비스 오류가 발생할 경우 데이터 전송이 불가하여 데이터 결측이 발생하는 원인이 된다. 그러므로 그림2와 같이 본부 RTDB를 제거하고 관리단에서 DB전송프로그램을 운영함으로써 관리단-본부 RTDB간에 발생하는 전송 오류를 줄일 수 있다.

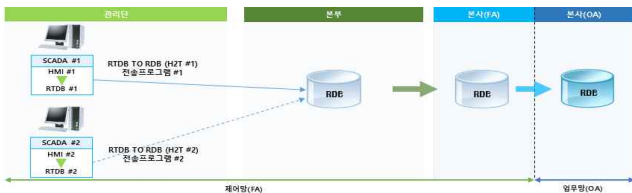


그림 2 데이터 전송체계 (1안)

나. 2안 : 전용 DB를 통한 데이터 전송

관리단 SCADA서버와 RTDB는 이중화되어 본부의 이중화된 응용서버로 서버~서버간 개별 통신이 가능한 구조로 되어있다. 그런데 관리단 SCADA서버를 유지관리하기 위해 RTDB를 정지시키는 경우, 데이터 결측을 막기 위해서는 먼저 전송 체계상의 절체가 정상적으로 진행되어야 한다. 그러나 본부에서는 관리단의 작업상황을 알 수 없기 때문에 서버 절체 등의 선행작업 없이 관리단에서 SCADA작업을 진행할 경우 데이터 결측이 발생하며, 이 경우가 데이터 결측 발생의 주요 원인 중 하나가 된다. 따라서, 그림3과 같이 관리단에 데이터베이스를 추가하여 관리단 SCADA서버의 데이터를 먼저 관리단 DB에 안정적으로 저장한 후 본부 DB로 전송하여 관리단 SCADA서버 작업에 의한 결측을 방지할 수 있다. 그러나 DB전송프로그램이 RTDB와 관계형DB(RDB)간 데이터 전송과 결측 복구기능을 모두 가지는 경우, 장기 결측데이터 복구의 어려움과 함께 서버에 부하를 발생시켜 서버 다운으로 인한 감시제어 불능으로 수도 사업장 운영에 심각한 문제를 초래할 가능성은 여전히 존재한다.

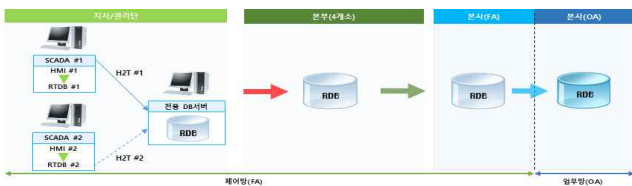


그림 3 데이터 전송체계 (2안)

다. 3안 : 데이터 전송 및 결측 복구기능 분리

제시하는 3안은 그림4와 같이 RTDB 상태감사·절체 및 결측복구 기능을 가진 별도의 프로그램을 도입하여 데이터전송과 결측복구 기능을 분리하는 것이다. 평소에는 RTDB 상태를 감시하면서 정상 운영 중인 RTDB에서 본부DB로 데이터를 전송하며, 그와 별개로 관리단DB는 RTDB의 데이터를 저장한다. RTDB 등에서 이상이 발생할 경우, 관리단DB로부터 본부DB로 복구프로그램을 통해 데이터를 전송하는 구조이다. 이 방안은 데이터전송 프로그램은 RTDB 데이터를 본부DB로 전송하는 기능을 담당하고 장기간 데이터의 결측복구는 별도 프로그램에 의해 관리단DB에서 본부DB로 데이터를 복구하여 서버의 부하를 경감할 수 있으며 데이터 전송상의 결측을 최소화할 수 있다.

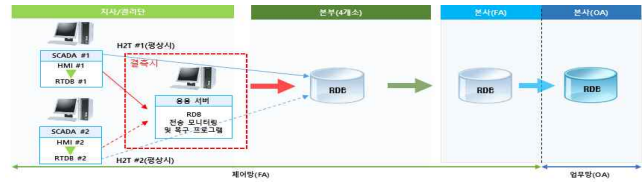


그림 4 데이터 전송체계 (3안)

2.3 최적 전송체계 제시

먼저, 2.1에서 데이터 결측의 원인으로 언급한 동기화 문제는 GPS 신호로부터 표준시간을 가지고 있는 동기화 서버와의 시간 동기화를 통해 공통적으로 문제를 해결할 수 있다. 2.2에서 제시한 방안들에 대해 장단점을 정리하면 다음과 같다.

구분	장점	단점
1안	<ul style="list-style-type: none"> 고장점 축소로 장애확률 감소 본부 관리(프로그램/서버) 경감 	<ul style="list-style-type: none"> 이중화 문제 존재 장기 결측데이터 복구 어려움
2안	<ul style="list-style-type: none"> 고장점 축소로 장애확률 감소 유지보수 등 지사 서버작업에 의한 본부 DB 결측문제 해결 	<ul style="list-style-type: none"> 추가 DB 등 투자비 발생 결측복구시 서버 부하가중 장기 결측데이터 복구 어려움
3안	<ul style="list-style-type: none"> 데이터전송과 복구작업을 분리하여 효율 및 안정성 향상 장기 결측데이터 복구문제 해결 	<ul style="list-style-type: none"> 추가 DB 등 투자비 발생 관리단 업무부담 가중

위와 같이 제시한 방안들은 각각의 특징 및 장단점을 가지고 있다. 데이터 결측문제 해결 관점에서 검토한 결과, 초기 투자비가 발생하지만 2.1에서 언급한 4가지 문제를 모두 해결할 수 있는 3안이 최적의 방안으로 판단된다. 별도로 IoT 단말기를 활용한 분기점 감시제어설비(TM/TC)의 데이터 결측을 방지하기 위한 방안을 아래와 같이 제시하고자 한다.

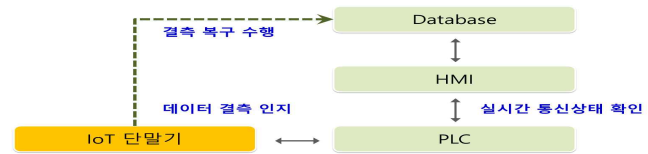


그림 5 분기점 데이터 결측해결 방안

III. 결론

본 논문에서는 수도사업장에서 양질의 데이터를 확보하기 위해 전송상 발생하는 데이터의 결측문제 해결 방안을 연구하였다. 수도사업장 데이터 전송 체계 현황 및 문제점을 분석하였으며 그 문제점을 해결할 수 있는 3가지 전송체계 개선방안을 제시하였다. 3가지 방안에 대한 장단점 분석을 통해 초기 투자 비용이 발생하지만 결측률을 최소화 할 수 있는 3안을 최적 방안으로 제시하였다. 본 연구가 데이터 결측율을 최소화함으로써 데이터 기반의 자율운영체계가 마련되기를 기대하며 향후 시범운영과 실증테스트를 통해 검증하고자 한다.

ACKNOWLEDGMENT

이 연구는 2020년도 산업통상자원부 광역협력권산업육성사업 연구비 지원에 의한 연구임(P0006174).

참 고 문 헌

- [1] 한국정보산업연합회, “제4차 산업혁명의 핵심은 데이터다”, FKII Issue Report 2016-06
- [2] 김종립 외 “데이터 고품질 관리를 위한 검정 알고리즘 최적화 연구”, 한국통신학회지, 2019
- [3] K-water “데이터 품질관리 지침”, 2018