

# 다중 센서 컨텍스트와 가변형 세로 압축을 활용한 산업용 게이트웨이에 대한 연구

변시우\*

\*안양대학교 소프트웨어학과

e-mail:swbyun@anyang.ac.kr

## A Study on Industrial Gateway using Multiple Sensor Context and Variable Column-Compression

Siwoo Byun\*

\*Dept. of Software, Anyang University

### 요약

본 논문에서는 열악한 산업 IoT 환경에서 불완전 특성을 분석하고, 이를 극복하기 위한 게이트웨이 모델과 처리 방법을 제시한다. 연결된 다중 센서의 복합적 의미 분석, 장애 징후에 대한 불안정성-메트릭 구성, 시간적 상호 연관성을 활용하며, 기초 논리 모델로서 복합 센서 컨텍스트 기반의 4계층 모델을 제안한다. 또한 센서 데이터는 수집 스트림의 흐름이 매우 긴 특성이 있고, 유사한 데이터가 연속되어 압축에 매우 유리하므로, 컬럼-기반 압축 기법을 제시한다.

### 1. 서론

산업 IoT는 건설 등의 열악한 산업 현장에서 빈번한 위험 징후를 실시간으로 예측하여, 큰 사고에 대처하는 산업 안전 융합 서비스이다. 그러나 고장 및 온도 변화에 의한 센서 감도 차이 등으로 인하여 신뢰도가 낮은 무의미한 데이터가 전송되거나, 아예 측정값 전송이 안 될 수도 있다. 특히, 배터리 및 저장 용량 문제로 데이터가 삭제될 수 있고, 업데이트, 리부팅에 의한 측정 중단, 또는 소프트웨어 버그에 의하여 오류가 전송될 수도 있다. 이는 결국 IoT 빅데이터 분석에 신뢰성과 효율을 저하시키며 통계 왜곡 및 오류를 유발시킨다[1-3].

최근 산업계에서 폰-노이만 구조에서 탈피하여, 메인 메모리에서 데이터를 바로 처리하여 효율성을 극도로 높인 인-메모리 데이터 기술이 점점 더 폭넓게 활용되고 있다. 주로 메모리를 대폭 증설한 중대형 서버에서 사용되며, 이제 소형 서버도 메모리 가격 하락으로 적은 용량이나 고속 처리 목적으로 특화된 적용이 필요하다.

최근 데이터 분석용 고속 저장소로서 세로(컬럼) 지향 저장소[4]가 확대되고 있다. 이 저장소는 기존의 한 레코드 단위로 연속 저장하는 가로방향-저장소와는 상반되어, 세로의 필드 단위(컬럼)로 분리, 저장, 검색하는 완전히 새로운 초고속 압축 저장소이며, 메모리 요구량이 많다. 메모리 가격 하락으로 적은 용량으로도 고속 처리가 가능하도록 IoT 소형 시스템에도 최적화, 경량화, 가속기법이 필요하다.

인-메모리 처리가 고속이긴 하지만, IoT용 소형 기기에는

램 메모리 용량의 한계가 너무 크므로 기존과는 다른 고압축 전략이 필요하다. 즉, 인-메모리 가속 기술과 고압축 세로-저장소 융합 시 수백의 고속 성능과 수십 배의 저장 효율 향상이 가능하다.

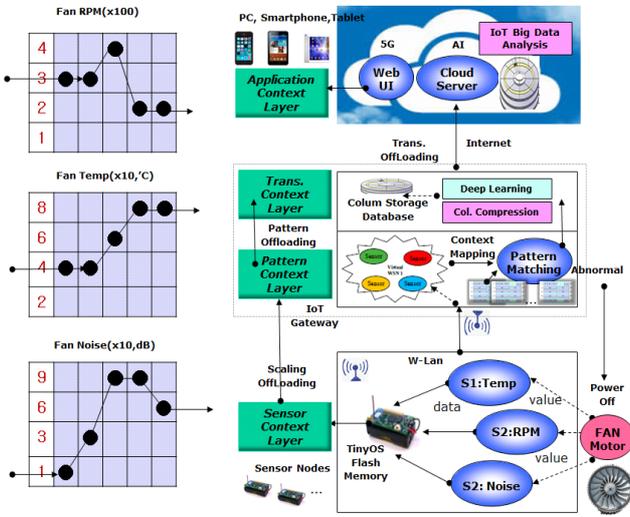
기본적으로, 이를 위한 매우 능동적인 오프로딩 기법[5]이 필수적이고, 복합적 센서-컨텍스트 활용, 컨텍스트-기반 필터링, 딥러닝 기술, 트래픽 셰이핑 등 연관 기술들이 요구된다.

### 2. 본론

#### 2.1 산업 IoT 환경의 불완전 특성 분석

기존 연구는 개별 특성 검출, 고장 예측, 특화된 질의, 패턴 필터링 등이 있었으나 열악한 산업 환경에 노출된 산업용 빅데이터도 아니었고, 주로 정제된 데이터만을 대상으로 하여 한계가 있었다. 따라서 불안정성에 중점을 둔 산업 IoT 데이터의 차별화된 구조 분석이 없었고, IoT 센서 고장 통계가 미흡하며, 오류/고장 프레임워크 조차 미흡하였다.

본 연구는 기존의 분리된 데이터 연구와는 다르게, 연결된 센서 의미를 중심으로, 장애 징후에 대한 불안정성-메트릭 (instability-metric) 맵이 필요하고, 시간적 상호 연관성, 센서 상호간의 영향력의 표현 모델, 설계 후 구축 및 정확성 검증까지 연계한다. 이를 달성하기 위한 기초 논리 모델로서, 그림과 같은 ‘복합 센서 컨텍스트 기반의 4계층 모델’을 제안한다. 이는 기존구조를 신속한 센서-컨텍스트 중심 및 저충단말



[그림 1] 복합 센서-컨텍스트 기반의 4계층 모델

중심의 선제적 오프로딩으로 차별화 한 것이다. 기존의 Analytical offloading model과 computational prediction scheme은 수식 연산에 의한 예측으로, 복잡하고 열악한 IoT 산업 환경에는 단순한 수식 계산이 어렵고 변화에 따른 융통성이 부족하다.

또한, adaptive (k+1) algorithm, negotiated resource contracts와 Partition components and automatic adaptation strategy가 있다. 여러 컴포넌트로 분해하고 자동으로 설정을 적응시키는 방법이다. 그러나 수십 가지의 다양한 센서 종류와 변화 추이가 변하는 산업 환경에서는 부족하다. 즉, 현장의 경험치와 복잡한 센서의 통합 패턴을 분석하고 인공지능 학습방법[6]이 필요하다. Execution history나 negotiated resource를 사용해도 adaptive (k+1) 알고리즘을 활용해도, 같은 이유로 미흡하다. Byte code transformer는 원격 호출을 사용하여 바이트 코드 변환을 하는 방식인데, 자바 환경에 효율적이고, Optimal program partitioning은 런 타임 오프로딩 변수를 실시간으로 추적하여 계산 모듈을 분산한다.

이 방식들은 코드나 계산 모듈 중심에는 효과가 있으나, 실제 적용하면 효율성이 낮다. 반면, 본 연구는 계산 코드가 아닌 데이터 워크로드를 분산하며, 오프로딩이 훨씬 더 방대한 실시간 IoT 데이터를 처리 가능하다. Load-balanced offloading scheme는 로드 밸런서를 통하여 전체적인 성능을 높이는 것이며, 데이터 자체의 양은 줄지 않는다. 반면에, 본 방식은 방대한 데이터를 초기에 줄이는 효과가 있다.

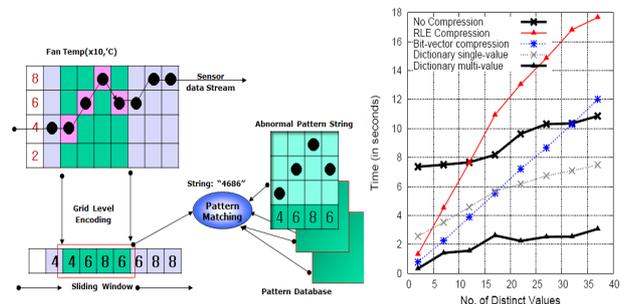
## 2.2 가변 트래픽 웨이핑과 세로-기반 선택 압축

산업 IoT 환경과 같이 불안정한 센서 노드로 구성된 불안전 무선 네트워크상에서는 자원 효율적인 데이터 통신을 위한 특별한 고려가 필요하다. 센서들은 24시간 주기와 같은 시간적 반복 패턴과 인접 노드간의 유사한 값이 발생하는 지역

적 패턴을 발생시킨다. 즉, 시공간적 로컬리티를 잘 활용하면 통신비용 절감이 가능하다. 또한, 산업 현장의 데이터 특성을 고려하여, ‘학습을 통한 스마트 스케일링’을 고려한다. 평상시에는 센서값이나 비디오 영상을 저해상도로 제공하다가, 공사 소음이나 진동이 큰 경우에만 고해상도로 센서 및 영상을 제공하는 것이다. 전송률 설정에 필요한 센서-컨텍스트들을 활용한 노이즈 센싱 예측(NSP) 알고리즘이 차별점이며, 이를 경험치 학습의 머신 러닝으로 통합한다. 더불어, 산업용 에지 디바이스의 컬럼-기반 데이터 압축, 고속 고장 진단을 위해 패턴 정렬을 활용한 속도 개선이 필요하다. 또한, 응답시간 지연 및 네트워크 부하를 더 줄이기 위하여 센서-컨텍스트 기반의 가변 비트레이트(VBR: Variable BitRate) 스트리밍 기술을 활용한다. 하지만, 주관적인 비트 정밀도의 단순 변환은 사용자가 느끼는 센서값의 질적 수준을 저하시킬 수 있으므로 사용자의 평가 결과를 추정할 수 있는 품질 측정 기법을 추가한다. 예를 들어, Structural SIMilarity (SSIM) index를 참고하되 SSIM의 많은 연산부하 때문에 실시간 환경에서 보다 적합한 도구를 적용한다.

## 2.3 시공-지역성을 활용한 컬럼 기반 선택 압축

센서 데이터는 수집 스트림의 흐름이 매우 긴 특성이 있고, 유사한 데이터가 연속되어 압축에 매우 유리하므로, 컬럼-기반 압축에 In-Memory 기술 융합으로 처리한다. 즉, 데이터의 특징을 잘 반영하면 고압축 및 메모리 가속 효과가 있고, 기간별, 요약 등의 관리도 용이하다. 특히, 센서 데이터는 읽기-집중 특성을 가지므로 읽기-연산을 최적화 하고자 하며, 통계 및 요약 데이터의 유지 관리도 고려한다. 본 실험실의 기상 데이터에 적용한 결과 80%이하로 압축이 가능하였고, 평균, 최저, 최대 등 부수적인 데이터를 쉽게 추가 가능하여, 불필요한 반복 연산을 줄일 수 있다고 판단된다. 또한, 센서-컨텍스트별로 반복적 파이프라인을 적용할 수 있는데, 이는 캐시용량이 크고, 멀티코어 추세의 CPU와 같은 최신 에지 컴퓨팅 환경에 잘 부합되므로, 병렬처리 효과를 극대화시킬 수 있다. 나아가서, 최근의 GPU는 수 백 개의 코어를 내장하고 있고 센서-데이터의 컨텍스트별 분류, 계층 분산, 분산 압축, 저장 및 검색이 많은 경우에 매우 유용하다.



[그림 2] 센서 스트림의 패턴 매칭 및 세로 기반 선택 압축 성능

또한 무선 센서 환경에서는 대부분 사고 등의 특정 상황을 제외한 정상시는 유사한 데이터가 계속 생성되므로, 유사한 그룹으로 분류하여, 전송 데이터를 압축함으로써, 전송량과 전송 빈도도 많이 줄일 수 있다. 센서 데이터 오프로딩을 위하여 스트림 패턴 분석 후 복합 센서 스트림의 패턴 매칭을 거치며, 유사 범위 컨텍스트의 판정은 슬라이딩 윈도우로 동적으로 변화되면서 유사 범위를 추적하는 기법을 활용한다.

### 3. 결론

산업 IoT는 열악한 산업 현장에서 위험 징후를 예측, 대비하는 융합 서비스이다. 본 논문에서는 산업 IoT 환경의 불안전 특성을 분석하고, 이를 극복하기 위한 게이트웨이 중심의 4계층 모델과 처리 방법을 제시하였다.

또한 센서 데이터의 시공간적 특성을 활용하여, 유사한 데이터를 선택 압축하는 컬럼-기반 압축 기법을 제시하였다. 다량의 센서 데이터를 초기에 선제적으로 간편하고 속도가 매우 빠른 패턴 매칭으로 선별하며, 복합 센서-컨텍스트를 이용하여, 상호 연관성을 조사한다. 이 때 연관성이 없는 무의미한 데이터를 대량으로 정제시키므로 효과적이다. 또한 그리드로 스케일링하여 축소 입력하면, 패턴 매칭과 데이터 저장 및 전송 속도를 크게 높일 수 있다.

### 참고문헌

- [1] K. Xiao, Z. Gao, Q. Wang and Y. Yang, "A Heuristic Algorithm Based on Resource Requirements Forecasting for Server Placement in Edge Computing," 2018 IEEE/ACM Symposium on Edge Computing (SEC), Seattle, WA, USA, 2018, pp. 354-355,
- [2] Ch.Sekhar, N.Suneetha. A Study on Data Categorization for Data Analytics, International Journal of Internet of Things and Big Data. Vol. 3. No. 1. May. 2018. pp:1-6.
- [3] Balaji Prabhu B V, Dakshayini M, An Efficient Data Analytics-Cloud Integrated Decision Support Service for Smart Agriculture, International Journal of Internet of Things and Big Data, Vol. 3. No. 12, 2018.01, pp.1-12.
- [4] S. Byun, and S. Jang. Asymmetric Index Management Scheme for High-capacity Compressed Databases, Journal of Korea Academia-Industrial, (2016), Vol. 17, No.7, pp.293-300.
- [5] Y. Son, J. Jeong and Y. Lee, An Adaptive Offloading Method for an IoT-Cloud Converged Virtual Machine System Using a Hybrid Deep Neural Network, Sustainability (2018), Vol.10, No.11, 1-15.
- [6] Sukhpal Singh and Inderveer Chana. QoS based Machine Learning Algorithms for Clustering of CloudWorkloads: A Review.International Journal of Cloud-Computing and Super-Computing. Vol. 3. No. 1. Jun. 2016. pp:13-24.