

# 공정 데이터 압축 성능평가를 위한 시뮬레이터 설계

유민형\*, 한상혁\*, 김영국\*  
\*충남대학교 컴퓨터공학과

e-mail : ymh0212@cnu.ac.kr

## Design of Process Data Compression Simulator for Performance Evaluation

Min-Hyung You\*, Sang-Hyuck Han\*, Young-Kuk Kim\*

\*Dept. of Computer Science & Engineering, Chungnam National University

### 요 약

산업분야에서 이용되는 공정 데이터는 데이터와 데이터 사이의 변화폭이 비교적 좁고, 많게는 수만 개의 포인트가 실시간으로 수집, 저장되어지는 대용량의 데이터라는 특징이 있다. 이러한 특성을 지닌 공정 데이터는 공정의 모든 데이터를 저장하지 않고, 전체를 대표하는 데이터의 일부만을 저장한다. 이러한 공정 데이터를 효과적으로 관리하고, 보관하기 위해 공정 데이터 압축 알고리즘에 대한 개발과 연구가 진행 중이다. 그러나 이미 알려져 있는 공정 데이터 압축 알고리즘과 개발하고 있는 알고리즘의 성능에 대한 평가를 할 수 있는 시뮬레이터는 개인이나 작은 집단만을 위해 만들어지거나, 공개되지 않고 있어 사용하기에 제한적이다. 이에 공정 데이터 압축 성능평가를 위한 시뮬레이터인 PDCS(Process Data Compression Simulator)를 제안하고자 한다. PDCS는 클라이언트, 서버기반으로 구성되어있고, 데이터 압축 알고리즘에 대한 성능평가가 가능한 시뮬레이터이다.

### 1. 서론

최근 컴퓨터 기술의 발전과 더불어 업무의 자동화를 촉진시켜 방대한 양의 데이터를 수집하고, 보관하는 것이 가능하게 되었다. 특히 산업분야에서 이용되는 공정 데이터는 데이터와 데이터 사이의 변화폭이 비교적 좁고, 많게는 수만 개의 포인트가 실시간으로 수집, 저장되어지는 대용량의 데이터라는 특징이 있다. 이러한 특성을 지닌 공정 데이터는 공정의 모든 데이터를 저장하지 않고, 전체를 대표하는 데이터의 일부만을 저장한다. 이러한 공정 데이터를 효과적으로 관리하고, 보관하기 위해 공정 데이터 압축은 필수적이고, 공정 데이터 압축을 위한 알고리즘 개발과 연구가 진행 중이다. 그러나 이미 알려져 있는 공정 데이터 압축 알고리즘과 개발하고 있는 알고리즘의 성능에 대한 평가를 할 수 있는 시뮬레이터는 개인이나 작은 집단만을 위해 만들어지거나, 공개되지 않고 있어 사용하기에 제한적이다[1-3].

이에 공정 데이터 압축 성능평가를 위한 시뮬레이터인 PDCS(Process Data Compression Simulator)를 구현하기 위한 설계를 한다. PDCS는 클라이언트, 서버기반으로 되

어있고, PDCS를 이용하여 이미 알려져 있는 데이터 압축 알고리즘에 대한 객관적인 평가를 할 수 있고, 또한 연구 중에 있는 압축 알고리즘을 비교, 분석할 수 있다.

본 논문의 주요 구성은 다음과 같다. 2장에서 공정 데이터 압축 알고리즘과 압축 알고리즘 성능평가를 위한 요소에 대해 기술하고, 3장에서 PDCS의 설계에 대해 자세히 기술한다. 4장에서는 결론 및 향후연구에 대해 기술한다.

### 2. 관련연구

#### 2.1 공정 데이터 압축 알고리즘

데이터 압축을 위한 알고리즘으로 BoxCar, Backward slope 등이 있다. 이러한 압축 기법은 특히 공정 데이터의 압축에 유용한 알고리즘으로 알려져있다.

BoxCar 알고리즘은 현재의 Data point와 과거에 마지막 기록되어진 Data point와 비교를 통해 데이터를 압축하는 기법이다. BoxCar 알고리즘은 이러한 두 개의 Data point가 Deviation값을 벗어난다면 현재의 Data

Point를 저장하는 알고리즘이다. Deviation은 데이터와 데이터 사이에 벗어난 정도를 의미한다. BoxCar 알고리즘은 변화의 폭이 상대적으로 좁은 공정 데이터의 압축에 유용하다. Backward Slope 알고리즘은 마지막에 저장되어진 두 개의 Data point를 이용하여 데이터를 압축하는 기법이다. Backward Slope는 과거에 마지막으로 기록되어진 두 개의 Data Point를 분석하여, recording limit에 대한 Deviation을 설정한다. 이 Deviation안에 속한다면, 이후의 Data Point는 저장되지 않고, Deviation의 밖에 Data Point가 있다면 현재의 Data Point를 저장하는 기법이다. Backward Slope 알고리즘은 BoxCar 알고리즘과 마찬가지로 변화의 폭이 좁은 공정 데이터의 압축에 유용한 알고리즘이다[4][5].

### 2.2 성능평가요소

압축 알고리즘의 성능평가 요소로는 Compression Ratio, Maximum Error, Average Error가 있다. Compression Ratio는 데이터의 원본과 압축 후의 데이터를 비교하여 압축률을 구한다. 이는 원본 데이터에서 데이터 포인트의 개수와 압축 후 데이터의 포인트 개수를 비교해봄으로써 알 수 있다. Maximum Error는 최대 에러율을 표현한다. 만약 원본 데이터의 어떤 포인트 값을  $v_1$ 이라 하고, 원본 데이터를 압축 후 decompression한 값을  $v_1'$ 이라고 한다면 에러율( $\epsilon$ )은  $|V_1' - V_1|$ 으로 계산될 것이다. 이와 같은 방법으로 최대에러율을 구할 수 있다. 또한 Average Error는 평균에러율을 의미하고, 각 에러율에 대한 합을 포인트의 개수로 나눔으로써 구할 수 있다. 이러한 압축 알고리즘의 성능평가요소는 PDCS의 결과데이터 속성에서 이용된다. 표 1은 압축 알고리즘의 성능평가요소를 나타낸 표이다[4][5].

<표 1> 압축 알고리즘 성능평가 요소

속성	설명
Compression Ratio	압축률
Maximum Error	최대에러율
Average Error	평균에러율

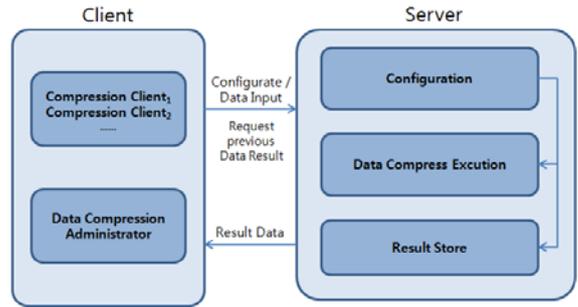
### 3. PDCS의 설계

3장에서는 클라이언트, 서버 기반으로 된 공정 데이터 압축을 위한 시뮬레이터인 PDCS(Process Data Compression Simulator)를 설계한다. 이 장에서는 PDCS의 구조, 설계원칙, 상세설계, 공정 데이터 압축 과정에 대

해 자세히 기술한다.

#### 3.1 PDCS의 구조

PDCS는 클라이언트와 서버로 구성되어 있다. 클라이언트는 Compression Client와 Data Compression Administrator로 구성되어 있고, 서버는 Configuration, Data Compression Execution, Result Store로 구성되어 있다. PDCS 구조의 상세한 설명은 3.3에서 기술하도록 한다.



(그림 1) PDCS의 구조

#### 3.2 PDCS의 설계원칙

<표 2> PDCS의 설계원칙

설계원칙	설명
확장성 (Extensibility)	알고리즘의 추가, 제거가 용이
신뢰성 (Reliability)	압축 전, 후의 데이터가 유사
유용성 (Usability)	사용의 편리성(GUI기반)

표 2는 PDCS의 설계원칙을 보여준다. PDCS는 확장성, 신뢰성, 유용성을 핵심 설계원칙으로 시뮬레이터를 설계하였다. PDCS는 공정 데이터 압축을 평가하기 위해 데이터 압축 알고리즘이 사용되는데, 이러한 알고리즘은 이미 공개되어진 알고리즘뿐만 아니라, 새로 개발되고 있는 알고리즘도 있다. 이에 PDCS는 데이터 압축 알고리즘을 라이브러리 형태로 추가, 제거할 수 있도록 객체 지향적 설계를 하였다. 이를 통해 PDCS는 확장성을 가진다. PDCS에서 압축 후의 데이터는 압축 전의 데이터를 간략하게 표현한 것이어야만 한다는 점을 설계원칙으로 삼고 있다. 가령, 압축 전과 압축 후의 데이터가 전혀 다른 결과를 나타낸다면 이는 신뢰할 수 없기 때문에 PDCS는 신뢰성을 설계원칙으로 설계한다. PDCS의 서버에서 압축 전과 압축 후의 데이터를 비교할 때, 에러율을 보여준다. 에러율은 압축 전의 원본 데이터와 압축 후에 다시 Decompression

을 거친 데이터를 비교함으로써 나타내어진다. 이러한 두 데이터 사이의 관계를 분석하여 결과를 나타냄으로써 신뢰성을 확보할 수 있다. PDCS의 클라이언트를 G.U.I(Graphic User Interface)기반으로 설계함으로써 사용자가 쉽게 이해하고, 사용할 수 있도록 설계한다. 또한 서버에 저장되어있는 알고리즘이나 구조, 압축 과정 등 관련된 지식을 사용자가 알지 못하더라도 이용할 수 있도록 추상적으로 설계한다.

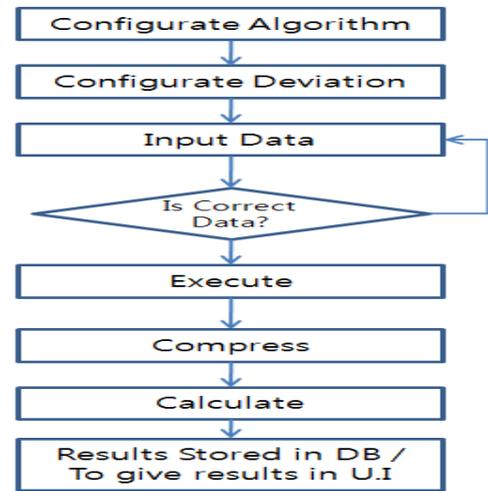
### 3.3 PDCS의 상세설계

클라이언트에는 압축을 요청하는 Compression Client와 결과를 통보받을 수 있는 Data Compression Administrator로 나누어진다. Compression Client는 PDCS에서 설정 가능한 알고리즘 선택과 데이터와 데이터 사이의 벗어난 정도를 의미하는 Deviation을 설정하고, 데이터 압축을 위한 공정 데이터를 입력을 한다. 이러한 설정값(알고리즘, Deviation)과 공정 데이터를 서버에 전송한다. Data Compression Administrator는 서버로부터 결과데이터를 받고, 이를 클라이언트에게 보여준다. 서버는 Configuration, Data Compression Execution, 그리고 Result Store로 구성된다. Configuration에는 클라이언트에서 입력한 설정값을 가지고 있고, 이를 전송받은 공정 데이터와 함께 Data Compression Execution으로 보낸다. Data Compression Execution에서 공정 데이터를 전송받으면, 압축이 가능한 데이터인지를 판단하게 된다. 압축이 가능하다면 Configuration값과 공정 데이터를 이용하여 압축을 실행한다. 데이터의 압축이 완료되면 공정 데이터의 원본과 압축 후의 공정 데이터를 비교하여, 데이터의 신뢰성 여부를 판단하고, 신뢰성이 확보되면 압축률, 최대에러율, 평균에러율을 계산한다. Result Store에는 원본 데이터와 압축 후의 데이터가 저장되어지고, 클라이언트에게 보여지기 위해서 결과데이터의 형태로 변환하여 클라이언트로 전송한다. 또한 각기 상이한 설정값으로 데이터를 압축할 때, 이를 비교할 수 있도록 테이블을 구성한다.

### 3.4 PDCS의 공정 데이터 압축 과정

그림 4는 PDCS의 압축과정을 도식화한 것이다. 가장 먼저 설정가능한 알고리즘을 선택하고, Deviation값을 설정한다. 설정을 완료한 후에 클라이언트는 데이터를 입력한다. 입력받은 데이터가 압축하기에 적당한 데이터인지의 여부를 판단하고, 압축이 불가능하다면 에러 메시지를 출력하고, 데이터를 재입력한다. 압축이 가능한 데이터이면 이를 실행시키고 서버의 Compressor는 선택되어진 알고리즘과 Deviation값, 원본 데이터를 입력받아 데이터를 압축하고, Calculator는 원본의 데이터와 압축된 데이터를 비교, 분석하여, 압축률, 최대에러율, 평균에러율의 결과를

도출한다. Calculator를 통해 얻어진 결과를 서버의 데이터베이스에 전송하면 데이터베이스는 결과데이터 속성에 맞게 포맷을 변경하여 저장하고, 이를 클라이언트에게 결과를 보여준다.



(그림2) PDCS의 공정 데이터 압축 과정

## 4. 결론 및 향후연구

본 논문에서는 공정 데이터 압축 알고리즘에 대한 평가가 가능한 공정 데이터 압축 시뮬레이터인 PDCS의 구조와 PDCS를 통한 공정 데이터 압축 과정, 결과데이터에 대한 속성들에 대해 기술하였다. PDCS를 통하여 이미 알려져있는 공정 데이터 압축 알고리즘과 개발 중인 공정 데이터 압축 알고리즘의 성능 및 에러율에 대한 객관적 평가를 내릴 수 있을 것으로 기대된다. 향후 연구에는 본 논문에서 제안한 PDCS의 구현을 통해 기존 공정 데이터 압축 알고리즘의 평가 결과를 도출한다.

### 참고문헌

- [1] S.Han, Y. Kim, "An Architecture of Real-time, Historical Database System for Industrial Process Control & Monitoring", CNSI 2011, 2011.05
- [2] 김운식, 모경주, 윤인섭 "클러스터링 기법을 이용한 공정 데이터의 압축 저장 기법에 관한 연구", Vol.4, No.4, KiGAS, 2000.
- [3] D. C. Barr "THE USE OF A DATA HISTORIAN TO EXTEND PLANT LIFE", Life Management of Power Plants, 1994.
- [4] M.A.A. Shoukat Choudhury, Sirish L. Shah, Nina F. Thornhill "Diagnosis of Process Nonlinearities and Valve Stiction: Data Driven Approaches", Springer, 2008.
- [5] 장대식 "수도통합운영시스템의 전송데이터 경량화에 관한 연구", 아주대학교 석사학위 논문, 2009.