Column-based Replication Control Scheme using Variable Compression and IoT Edge Gateway

SiwooByun[1]*

*1 Professor, Department of Software Engineering, Anyang University, 22 Samduk-ro, 37th street, Anyang-shi,
708-113,South Korea
swbyun@anyang.ac.kr[1*]
Corresponding author*: mobile Phone:+82-010-2776-0922

Article Info
Page Number: 456 – 463
Publication Issue:
Vol. 71 No. 3 (2022)

Abstract

IoT edge computing reduces cloud computing's overload that redirect local data to a central data center. This research proposes a new reliable data management scheme using IoT edge storage.Because IoT sensor network uses unstable wireless media and narrow bandwidth, IoT applications could suffer from unreliable and ill-timed data services. IoT data replication canimprove data availability because each sensor node can use its own data. Therefore, efficient replication scheme is much more important in sensor network environment rather than in typical distributed environment.

The purpose of using sensors is to check if the desired devices is functioning normally within acceptable range or if abnormal events have occurred. Therefore, it is not necessary to deliver too much detailed values to the server if they are within the allowable range. It is much more efficient to break down the overall range of sensor values into different interval units. Proposedvariable scaling method can reduce the amount of data communication and storage capacity by replicating only the shorten code of the multi-scale area rather than fixed scaling depending on the context of the sensor data.

The differences between proposed replication schemeand previous schemeare as follows. First, proposed schemecopies the replicated data inside the remote IoT edge. Second, the sensor data being replicated is a column-based compressed version, not the same as the original. Third, the replicated sensor data is stored at a lower resolution than the original.Experiments show that the size of the sensor data can be minimizedto14%.Since proposed scheme can supportefficientdata servicefor unstable IoT environments, it can be used for reliable sensor monitoring applications.

Keywords— IoT database, sensor data scaling, column-based compression, data replication, edge computing

Article History
Article Received: *12 January 2022*
Revised: *25 February 2022*
Accepted: *20 April 2022*
Publication: *09 June 2022*

1. Introduction

IoT(Internet of Things) sensor network has received significant attention in smart system areas. Small IoT devices can be designed with on-board calculations, wireless communications and sensor detection abilities(Figure 1).Each information device could include its tiny managementtools and a small database. Recent work has also begun exploring the potential applications for measuring various IoT environments[1,2,3].

Figure 1. Examples of Tiny Sensor Nodes

IoT applications include energy usage monitoring and planning energy conservation in buildings, military and private surveillance, natural habitats monitoring for understanding environmental dynamics, and collecting data for learning environments.IoT sensor network is different from the traditional stable networks since IoT applications automatically operate unattended and IoT sensor devices use limited battery and narrow wireless channel.Most IoT applications are data-centric since sensor nodes are designed from the point of measured data rather than identified data such as ip address of conventional networks. That is, measured data is most important in IoT networks and sensor network is treated as a huge database called *sensordatabase*[4,5,6].

2. Backgrounds

2.1 IoT Dataand Storage Model

IoT sensor devices have the following resource constraints. First, the wireless network connecting the sensor devices provides insufficient quality of data service, has limited networkbandwidth, frequently drop packets, and network latency with high variance. Second, sensor devices have limited battery power. Third, IoTdevices have limitedcomputing memory andCPU power [6].

Recent column-based storage model [7,8] is more advantageous than general storage models for storing sensor data(Figure 2). A column-based data storage store data in the order of columns and not in the order of rows(records) as in general data storages. Traditional row-based storage systems have to read many unrelated data from data repository, so they have performance limitations when handling a number of concurrent, diverse queries. However, in column storage model, only the columns related to IoT query need to be taken from the data repository, which is densely compressed to improve read efficiency like the column-store[8].

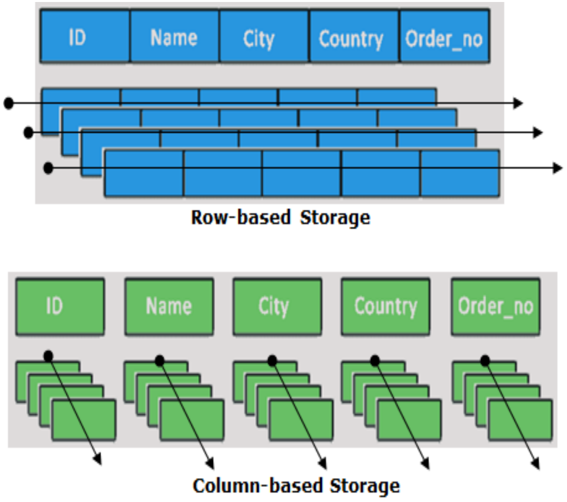

Figure2. Column-based Storage Model

2.2 IoT Edge Computing

IoT edge is an endpoint device that is used by people or is built around us. IoT edge computing[9,10,11] is defined as a distributed computing infrastructure that includes a large number of sensor devices that are well connected to each other. It is a new paradigm that provides high computing resources close to distributed IoT devices. Therefore, IoT edge node can collectand analyze raw sensor streams locally instead of transferring them to the fog server or remote cloud server, significantly reducing traffic overheads and speeding up the analytic task. However, where IoT edge nodes should be placed to facilitate communication is important (Figure 3).

Edge computingtechnology reduces cloud(fog) computing's overload that redirect local data to a central data center. In the future, special AI chips with better processing power, storage and other advanced features will be installed in various edge devices. As 5G technology matures in the future, the expanded edge devices will establish a more robust communication line with central services. This is because 5G provides lower latency and higher bandwidth. In the future, IoT will lead edge utilization, and most of the work will be done on edges rather than fog or centralized clouds.

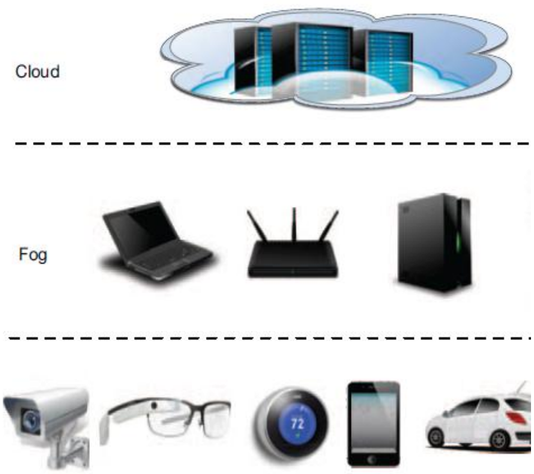

Figure 3. IoT Edge-based Computing with Small IoT Sensor Devices

2.3 Reliable Management for IoT sensor data

Because IoT sensor network has unstable wireless media, IoT applications could suffer from unreliable and ill-timed services. One way that reduces possibility of the undesirable services is data replication ineach sensor nodes. In data replication environments, each IoT sensor node can use its own data in case sensor node failures or wireless network failures continue for a long time. Moreover, the operation cost of wireless communication is very expensive as compared to the other operations. As compared to general environments with sufficient bandwidth and reliable power, data transmission overhead should be minimized for actual implementation of data replicationfor unstable IoT sensor networks.

3. Sensor-aware Variable Replication Scheme using IoT Edge

3.1 Replication management using IoT edge device

In this study, proposed schemeaims to offer high levels of reliability and scalabilityas well as fast response. Advantages of proposed edge computing model include a significant reduction in data transfer and reduced congestion.The proposedscheme is called *SaVR* (Sensor-aware Variable Replication) that improves general mirroring techniques to enhance storage performance and reliability of sensor data managementin IoT edge environment(Figure 4). Several RAID levels were considered before our scheme design, RAID1 called mirroring is determined to be a primary design base, because higher level RAID schemes induce complex computational overheads such as data parity and distribution to small-size IoT edge device.Proposed technique keeps sensor data in the same way as RAID1. However, instead of simply copying the original data, the variably scaled sensor data is written to the edge storage after performing column compression.

The differences between proposed technique and normal mirroring are as follows. First, this technique copies the replica inside theremote IoT edge connected to the local network, whereas normal mirroring copies it inside its sensor node. Second, the sensor data being replicated is a compressed version, not the same as the original. Third, the replicated sensor data is stored at a lower resolution than the original, taking into account the characteristics of the sensor. Fourth, the replication of the sensor is not performed immediately, but the reflection cycle varies depending on the context of the sensor data generation.

For example, sensor data is generated by sensors every second at a sensor node. However, it is aggregated every minute at the edge gateway, and it could be sufficient to upload the aggregated data to the cloud server on an hourly basis. This is because only the demanded data should be passed and aggregated through the IoT edge device to reduce network transmission overhead in the first filtering stage. In addition, the data required by the cloud server could be removed or refined for the same reason.

The first reason why it needs to be mirrored on the IoT edge is due to the instability of the sensor node. The second reason is that the volume of sensor data is too huge to be mirrored in the cloud sever. Thus, the IoT edge should reduce the data volume to a transferable size for outbound internet by filtering sensor data. Third, IoT edge device is fully connected to short-distance LAN with sensor nodes, so they have relatively low communication workload and are stable LAN. Fourth, compared to tiny sensor nodes, IoT edge gateway can provide much more stable hardware and power and is not exposed to the outside world damage.

However, effective storage techniques are important since the amount of mirrored sensor data should be reduced sufficiently to be processed even under the resource-poor edge device as compared to the average computer.

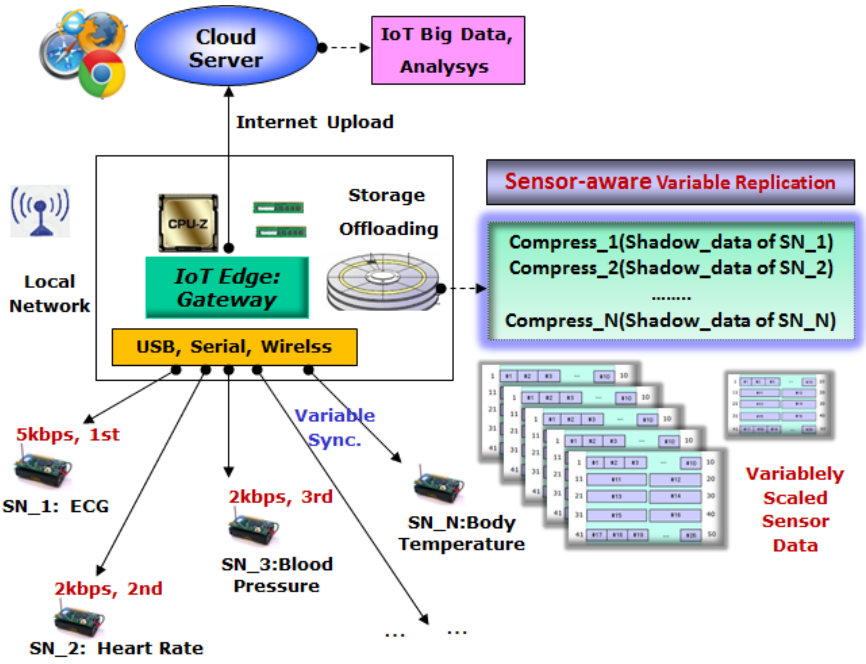

Figure 4. Concept of Sensor-aware Variable Replication Model

3.2 Variable scaling and compressed transmission using sensor context

The purpose of using a sensor is mostly to check if the desired devices is operating normally within acceptable range or if abnormalities have occurred outside of it. Therefore, it is not necessary to send too much detailed values to the server if they are within the allowable range. Usually, it is important for servers to have average statistics of many sensor values or sensor values that have abnormalities in each region. That is, if the approximate mean value is required rather than each sensor data or accurate sensor data is not required until theabnormalevent, scaled data compression is more suitable.

On this point, it is much more efficient to break down the overall range of sensor values into different interval units, rather than dividing them by the same interval. That is, the variable scaling method can significantly reduce amount of data communication and storage capacity by transmitting only the short code of the multi-scale area rather than fixed scaling method according to the context of the sensor data.Of course, mathematical, computational, and statistical methods in existing multi-scale data studies [12] have been used to effectively describe for the occurrence of events(Figure5). This study expands multi-scale model, and uses variable scale according to the characteristics of sensor data, and mirrors data based on column-compression.

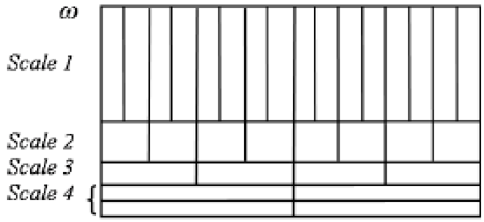

Figure 5.Concepts of Multi scaling Model

For simple example, assuming that a certain workshop has 100 same sensors, that the normal temperature range of the sensor is 10 to 40 degrees, that it is safe to report in 5 degrees at the normal range, and that below or above this range is an alarm target and should be reported in 1 degree increments. The overall sensing range is 1 to 50 degrees.

For this case, the variable scale set (*VS_1*) can be expressed in order of (#area code, temperature), and the whole area can be presented as follows (Figure6). In this case, the data can be reduced to 26/50 by almost half.

VS_1= { (#1,1), (#2,2), ..., (#10,10),
(#11,11~15), (#12,16~20), (#13,21-24), (#14,25~30), (#15,31~35), (#16,36~40),
(#17,41),(#18,42), ..., (#26,50) }

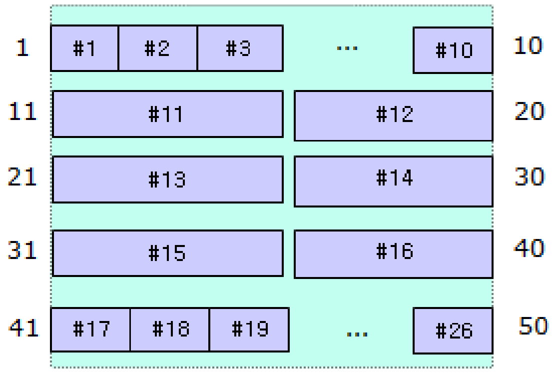

Figure6. Concepts of Sensor-aware Variable Scaling Model

In addition, it is possible to minimize the amount of the data transfer by using variable cycle transfer. For example, if a sensor node sends data in the normal range every 3 seconds and sends data in the abnormal range every 1 second, it can reduce the amount of sending data by one third. In fact, most of the data is normal, so this may reduce it further. In addition, suppressing repetitive similar data can enhance performance much more.

3.3 Scaled CompressionTest for ReplicatedSensor Data

Performance test was conducted to measure the variablescalingand column-based compression effects of proposedSaVR. Test resultsare describedin Table 1 and showthe size comparisons of original sensor data andresultsensor data. General sensordata can be compressed toabout 40% by numerical repetition and token mapping skills.

Table 1 shows the performance of scaled compression scheme is satisfactory.In Table 1, the size of the original sensor datawerereducedto4~14% because SaVRminimizedthe replication overhead bysensor-aware scaling and column-based compression unlike the traditional datamanagementtechnique.

The column-based compression modules based on lzo-API[13] are constructed by multiple chain connections with small compression segments rather than full block compression, in order to reduce both compression and decompression overhead as much as possible.

Table 1. Results of Scaled Compression Test

Sensor Database (Col_No. : Attribute)	C1: Battery Level(%)	C2: Outside Temp.(’C)	C3: Relative Humidity(%)	C4: Switch#1 (On/Off)	C5: … Wind Speed (m/sec)
Example values	25%	22.3’C	61%	T/F	8m/sec
Original data size (1000 records, Byte)	4,090	4,268	3,231	3,210	3,702
Scaling ratio	1:2	1:3	1:2	No	1:4
Data size after test (1000 records, Byte)	→354	→624	→408	→84	→188
Scaled-compression result (100→%)	→8%	→14%	→12%	→2%	→4%

4. Conclusions

In this study, we focused onunstable sensor environments, and analyzed IoT edge computing technology. In addition, we proposed a replicated data management scheme using IoT edge, named *SaVR*, in order to cope with unstable sensor nodes in IoT sensor networks. Unlike the previous approaches, SaVRscheme exploits a sensor-aware variable scaling technique and column-based compressiontechnique to improve datareliability and replication efficiency in sensor data collection and motoring. By applying scaled-compression skill, SaVRcan store replicated sensor data into IoT edge storage to achieve the space efficiency and reliability.

Experiments were conducted to measure the compression and scaling effects of SaVR, and the results showthat the size of the original sensor data can be minimizedto14% at least.This is because SaVR successfully reducesthe replication overhead through sensor-aware scaling and column-based compression as compared to traditional row-based database schemes.The future study includes a secure replicationcontrol, reliability model forbig sensordata,unstableIoT data analysisand the medical edge computing.

5.References

1. BonnetP., GehrkeJ., SeshadriP. Towards Sensor Database Systems, Proceedings of the Second International Conference on Mobile Data Management,2011 Jan; 3-14, Hong Kong, 1-12.
2. Hye Yun Kim, Seong Cheol Kim, Hyun Joo Park. Priority and Delay Aware packet transmission MAC Protocol for Wireless Sensor Networks. International Journal of Security Technology for Smart Device. 2018; 5(2):9-14.
3. Intanagonwiwat C., Govindan R., Estrin D. A Scalable and Robust Communication Paradigm for Sensor Networks. In Proceedings of the Sixth Annual ACM/IEEE International Conference on Mobile Computing and Networking, 2000 August 06-11; Boston, Massachusetts, USA, 1-12
4. YaoY., GehrkeJ. Query Processing for Sensor Networks. IEEE Pervasive Computing. 2003;3(1):46-55.
5. Lala, A., Bhaskar, A., Chakrabarti, P. Review on energy efficient approaches on dynamic source routing protocol in MANET. International Journal of Smart Device and Appliance. 2017; 5(1):1-10
6. Siwoo B., Dynamicsensordatamanagement scheme for IOT servers, Journal of Critical Reviews ,2020; 7(14):396-401
7. Ahn S., and Kim K. A Join Technique to Improve the Performance of Star Schema Queries in Column-Oriented Databases, Journal of Korean Institute of Information Scientist and Engineers. 2013; 40(3):209-218.
8. Abadi D., Samuel R., N. Madden, ColumnStores vs. RowStores: How Different Are They Really?, Proceedings of the ACM SIGMOD'08, 2008 Vancouver, BC, Canada, 967-980.
9. Jie-Won J., Young-Mo K. An Empirical Study on the Adoption Intention of IoT Based Smart Gas Safety Shutoff Device Service. International Journal of Internet of Things and its Applications. 2018;2(2):7-12.
10. Ramanpreetv Kaur Deol, Jinan Fiaidhi, Sabah Mohammed. Intruder Detection System Using Face Recognition for Home Security IoT Applications: A Python Raspberry Pi 3 Case Study. International Journal of Security Technology for Smart Device. 2018;5(2):21-32.
11. GopikaPremsankar, Mario Di Francesco, Tarik Taleb, Edge Computing for the Internet of Things: A Case Study. IEEE Internet Of Things Journal. 2018;5(2):1275-1284.
12. Marco S. Reis. Multiscale and Multi-Granularity Process Analytics:A Review. Processes. 2019 Jan.; 7(2):61;1-21,https://doi.org/10.3390/pr7020061
13. Professional data compression [internet],Lzo; 2020 [updated 2017Mar.01; cited 2020Aug.19].Available from: http://www.oberhumer.com/products/lzo-professional/ (website)