Data Prioritization at Multi-user IoT Gateway with Multiple Sensor Data Attributes

Sunyanan Choochotkaew[1,a)] Hirozumi Yamaguchi[1,b)] Teruo Higashino[1,c)]
Megumi Shibuya[2,d)]

Abstract: Nowadays, use of sensor-derived information for obtaining the significant knowledge in smart applications is rapidly increasing in several contexts. One of the most popular problems is data sending from a sensor gateway to a cloud service. It is likely that remote monitoring sites such as houses with elderly people in rural areas are connected via resource-limited, unstable 2G/3G networks. Moreover, even in urban areas, it is desirable to reduce the volume of sensor data based on the contexts of applications since each sensor usually continues to generate data regardless of the application contexts and correlation among similar data. Sending all such data is quite inefficient as most of the data is discarded without being utilized at the server. Considering those requirements to reduce the data volume maximizing the satisfaction of requirements by applications, in this paper, we propose a sensor data prioritization approach. The approach utilizes an MCDA (Multi Criteria Decision Analysis) technique called REMBRANDT to aggregate requirements and to determine a single priority value over sensor data with different attributes. Through a building management system case study, we have shown that our method can obtain higher satisfaction than naive transmission strategies in practical situations.

1. Introduction

The *Internet of Things* is broadly utilized in a variety of application such as health monitoring, product monitoring, structural monitoring, smart appliance control and smart buildings [9], [10], [11], [15], [16], [18]. Most of the above-mentioned applications can be driven by up to petabyte scale of sensor data. However, not only high volume of the collected values and their metadata, sensor data also have variety and velocity [12]. As such a variety of data is continuously generated from different types of sensors, data should be delivered to a cloud service and be analyzed for detecting some events. However, it is often likely that the capacity of communication link between a gateway at monitoring field and a faraway server is limited. For example, from the previous work [17], the real-world system that the multi-application server remotely controls the branch building in a rural area (near Toyota city) have been developed and deployed. In the building, 3G network is available but it is not stable due to remote distance between base stations and the mountainous area and causes some losses of data.

Several researchers consider this link problem and try to work out at the gateway side. Many of them focus on compression techniques [1]. However, there are not so many studies that consider the "importance of values" according to the requirements of applications and select data to be prioritized to meet the requirements. We concern the general fact that among the large amount of data, not all but just some data is important. For example, in a lighting control system of smart buildings, the data from movement and occupancy sensors are mainly used while those from temperature and heat detectors have no any effect. In a smoke detecting system, the data of smoke and gas detectors which are higher than the safety threshold are much more important than the data with normal values. Bisdikian et al. have described the concept of value assessment for WSNs in term of *Value of Information* (VoI) [2]. VoI is an "assessment" of the utility of information (a kind of utility function) derived from *Quality of Information* (QoI), which is the innate information property. In other words, QoI is the original characteristics of sensor data such as accuracy and coverage while VoI is the attribute which is determined later by QoI and usage context for ascribing sensor data. For example, thematic relevance of the movement and occupancy data will be assessed as high in lightening-control context but low in smoke-detecting context. According to the assessed value, we can numerically show that how different the importance of the data from movement and occupancy sensors for lightening control system and the importance of that for smoke detecting system. [2] has also presented a framework to determine VoI based on a multi-attribute decision-making technique.

However, there is still an important issue to be concerned, which has not been considered in the past literature. As stated earlier, different sensors are used by different applications. In addition, those applications may have different priorities among

1 Graduation School of Information Science and Technology Osaka University, Osaka, Japan
2 KDDI R&D Laboratories, Saitama, Japan
a) sunya-ch@ist.osaka-u.ac.jp
b) h-yamagu@ist.osaka-u.ac.jp
c) higashino@ist.osaka-u.ac.jp
d) shibuya@kddilabs.jp

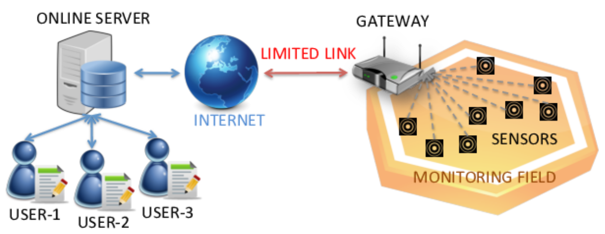

Fig. 1 Multi-user IoT Gateway

themselves and have a wide variety of requirements. For example, some require timeliness of data, and some others need high spatial resolution of sensor data instead of timeliness. Therefore, it is desirable that the utility functions can explicitly consider both application priorities and importance of values (VoI) in a best-mixed way, and a framework is necessary to enable such functions.

In this paper, we design a data prioritization system running on a gateway of wireless sensor network (WSN) consisting of IoT devices. The WSN may be deployed at a smart house, smart building, or any other facility or area to monitor a variety of values such as temperature, humidity and human location. The gateway bridges the WSN and a remote server where different remote applications (*i.e.* service providers) with different priorities may issue different types of data requests over sensor data with multiple attributes. Similar to [2], we take one of the Multi-Criteria Decision Analysis (MCDA) techniques for data-value evaluation. However, we consider the following scenario where different applications (or users) may require the data with different attributes (Fig. 1). In such a scenario, the gateway needs to aggregate all the requirements and properties of data and to derive unified priorities to be used to determine the priorities in the data queue of the gateway, which is connected to the remote server via public networks. To do so, we introduce the REMBRANDT method [8] to assess the value of data based on requirements, while preventing undesired phenomenon called rank reversal [5]. The assessed value will be used as a condition for priority queue of the gateway. In case that buffer queue exceeds the limit, the lowest priority data will be thrown away.

We have introduced an example monitoring system as a practical case study, and the observed knowledge is applicable to our real data collection system presented in [17]. We have shown that our system can achieve the best satisfaction to the given requirements, among other prioritization policies.

Moreover, we also concern the practical case of the multi-user system. Generally, each user can have a different kind of requirements of sensor-data value. In addition, users among themselves may also have different levels of priority as well. For example, in the smart building system, users who are responsible for controlling switching device always have lower priority than the ones who require data for security and safety application. So, for supporting the multi-user case, we also proposed the process for merging all user requirements into just one merged requirement with consideration of value definition in each requirement and the different priorities between users.

The rest of paper is organized as follow. Section 2 is about the related work which mentions the existing compression solutions and data prioritization. In this section, our contributions will be settled as well. Our proposed system is introduced in Section 3 along with an illustrative example. Based on the example, we set up the environment for simulation and show the results of the performance evaluation in Section 4. Then several issues are discussed for future research direction in Section 5.

2. Related Work

To tackle the challenge of transferring the large amount of sensor data over capacity-limited links, many efforts have been dedicated so far. Each of those existing solutions can be implemented at the origin like a gateway or sensors, in a middle medium, or at the destination like a server. We focus on the first one, which can further be categorized into two types. The first one is optimization of data collecting, which is implemented at the sensors. The second one is optimization of data sending, which is done at a gateway. We are interested in the second type. Most of the approaches on this type utilize compression techniques. In this paper, we consider a different approach that can additionally be used with those compression-based methods. Due to the variation of data values, we are more focusing on data prioritization. In the followings, we will discuss the related work in detail.

2.1 Compression Solution

This kind of solution is to compress the data into small size by encoding before sending and decoding back at the destination. The efficiency can be evaluated in terms of size, time, and loss of contents. One of known implementations of such compression techniques is Packedobjects (PO), which is a compression library for XML with easiness of implementation, high reliability, and acceptable efficiency. According to [1], with using PO library, the data size will be reduced to only 8% of the original size. The size of a POcompressed-XML sensor value is about 31 bytes. Moreover, the effect of latency and packet loss is significantly low and the time of decompression is fairly low.

2.2 Data Prioritization in WSN

There are some data-centric communication systems which manage resources by concerning the context of data [3], [6]. However, all data in the same stream are still fairly treated even though their VoI are totally different. Instead of that, we would like to use available resources, such as bandwidth and buffer size, more properly concerned with data's VoI. First of all, we must assess the VoI of data referencing to user requirement. An existing sensor-value-concerned paper [2] adopt one of the Multi-Criteria Decision Analysis (MCDA) techniques called Analytic Hierarchy Process (AHP) [4] into their work for value assessment. Unfortunately, AHP confronts with a heavy criticism of undesired phenomenon called rank reversal [5]. Later, the modified version of AHP called REMBRANDT has been proposed in [8] for preventing that phenomenon. So, we took this REMBRANDT method into consideration for assessing sensor-data value in our system at the first step. Then, the assessed value will be used as a condition for priority queue of the gateway. In case that buffer queue exceeds the limit, the lowest priority data will be thrown away.

According to the definition in [2], Value of Information (VoI) is an assessment of the utility of an information product when used in a specific usage context. The authors of [2] have also defined the taxonomy of VoI attribute in sensor networks and present how VoI depends on the quality characteristics of information (QoI) with mentioning easily-derivable relations from QoI to VoI. For a complex value such as information relevance, some researches particularly study assessing functions [13], [14] which are out of our scope. Though most of those approaches have focused on the selection of sensor data with multiple criteria, there is no or just a few research that apply the knowledge to the sending process at the gateway like our proposed method.

2.3 Our Contributions

The main contribution of our work is to introduce the new fashion of efficiency improvement dealing with the Value of Information of sensor data. We exploit a MCDA technique for determining priority of each sensor data on Priority queuing feature of the congestion management (QoS) at the sensor gateway. We propose a fast and flexible approach of score assignment in prioritization process by pre-calculating approach. We modify the original fixed-alternative REMBRANDT algorithm to multiple generalized components (i.e. independent alternatives), and cache the weight-per-alternative for each criterion. We also propose a merging algorithm for multi-user support with high similarity and correlation representing all user requests. Through simulation experiments, we have shown the efficacy of our method in terms of satisfaction values.

3. Proposed Method

3.1 System Architecture

As depicted in Fig. 1, our scenario is separated into two sides, gateway side and online-server side, which are linked by the Internet. An online server is stationed in urban area while a gateway and sensors are placed in the low-resource monitor field. The server-attached resources are abundant while the gateway-attached resources are limited. The limited resources in our consideration mean low bandwidth of medium and small size of gateway buffer. Multiple applications (or users) may request different kinds of sensor data from the server and each user has a different priority. We note that our study does not involve the data collection process from sensors as well as the query process at the server.

The gateway in our system prioritizes the sensor data with its value before sending. We introduce a distinctive concept to improve the sending process of sensor data over capacity-limited links. Our system is developed under the practically possible assumption that we can differentiate the value of each sensor data. We basically apply the REMBRANDT algorithm (explained in the next subsection) for value assessment. However, we extend a concept of "alternatives" (*i.e.* sensor value attributes). Instead of fixed alternatives for all criteria, our criteria can have different sub-alternatives, and the final alternative is a combination of one of the sub-alternatives from all criteria. This allows more flexible specification of priority requirement from each application. The main process of the selection system is the *value-based selection process* and the supplementary process for multi-application scenarios is the *request-merging process*.

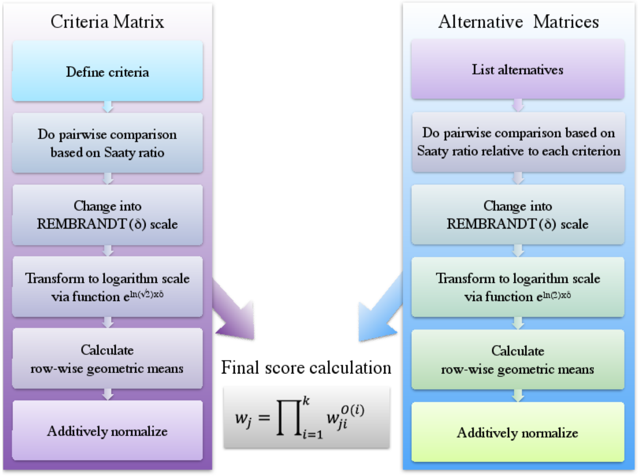

Fig. 2 Summary of Rembrandt's process

3.2 Multi-Criteria Decision Analysis and REMBRANDT System

According to [4], *Multi-Criteria Decision Analysis* (MCDA in short) is an analysis for making a decision among existing alternatives against the considered criteria. MCDA techniques are developed for handling large amounts of complicated information in a consistent way. They are used for many purposes such as identifying the most preferred option, ranking, bounding list, and distinguishing acceptable possibilities. Related to our study, among a number of techniques, we are interested in an Analytical Hierarchy Process (AHP) technique. The AHP technique derives weights and scores based on pairwise comparison between criteria and between options. We adopt the basic concept of this technique because pairwise comparisons are easier to determine and generally accepted. However, there are some criticisms about the original AHP devised by Saaty [5]. The most significant phenomenon is called *Rank Reversal* [5], [8]. There are many methods that modify the original one to avoid or get rid of this phenomenon. According to [8], the most effective one is the *REMBRANDT* system.

The significant feature of a REMBRANDT system is a multiplicative version of AHP approach. It uses geometric mean in the relative-value calculation instead of the arithmetic mean to prevent rank reversal. The calculation can be summarized as depicted in Fig. 2. The first step is to change members in comparison matrices to REMBRANDT (δ_{jk}) scale expressed as an exponential function of the difference between the echelons of value on geometric scale defined by Lootsma in [7]. The next step is weight calculation. The REMBRANDT-scale members are further transformed into logarithm scale as an exponential function of itself multiplied by scale parameter using ln $\sqrt{2}$ for criteria and ln2 for alternatives. Then, row-wise geometric means are trans-

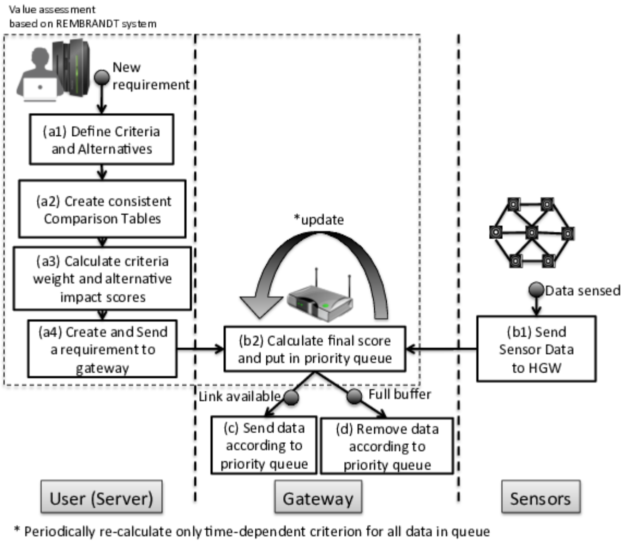

Fig. 3 Value-based selection process

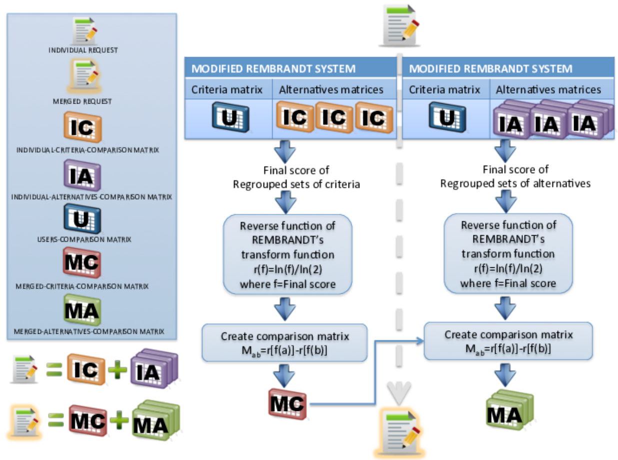

Fig. 4 Request-merging process

formed into values called weight. The last step is to aggregate the final score. Since it is a multiplicative method, the final score of each alternative (w_j) is calculated by product function as in (1),where w_{ij} is the weight of alternative j of criteria i and $O(i)$ is the normalized weight of criteria i. The benefits of this method is proven in [5].

$$w_j = \prod_{i=1}^{k} w_{ji}^{O(i)} \tag{1}$$

3.3 Value-based Selection Process

There are four activating events related to this main process as shown in Fig. 3. The first event is that an online server gets a new request from an application. Such an application must define *criteria* and corresponding *alternatives*, and then define *consistent comparison tables* of them. Then the online server will create a *requirement* according to the request to be sent to the gateway. The requirement includes definitions for classifying alternatives in each criterion and a *reference score table*. The reference score table contains a pre-calculated impact score, which is a row-wise geometric mean powered by row-wise geometric means of its parent criterion, pending at the weight calculation step of the REMBRANDT system. The second activating event is that sensors send their collected data to the gateway. After that, the gateway classifies the data into one of the defined alternatives for each criterion and continues the last step of calculation. Then the gateway uses the results for prioritizing data in its sending queue by using priority queueing feature of the congestion management. If there is a time-dependent criterion, the value is recalculated and the queue is resorted periodically. According to priorities in the sending queue, the gateway sends higher-priority data before the lower one when a link is available for data transmission, and removes the lowest ones when a buffer is full.

3.4 Request-merging Process

Request-merging process is a supplementary process for multi-application support. In a multi-application system, there is more than one different request from applications in step (a1-a2) of the main process (see Fig. 3). Before continuing the rest of main process, the multiple requests must be merged into one request first. The merging process is performed as depicted in Fig. 4. The first step is calculating the final impact score of regrouped sets by our modified REMBRANDT algorithm which supports flexible alternatives. We start with finding the regrouped sets of all requests in criteria and alternatives. The regrouped sets consist of intersection sets between any relative sets and the rest of subtraction sets. Note that the number of members in regrouped sets of criteria is always one. Slightly different from the criteria, the alternative has two types: set and range. Set-type alternatives can be handled exactly in the same way as the criteria. For range-type alternatives, we determine one continuous range as one set. Most operations are similarly conducted for both types of attribute. Except for minus operation, the leftover part from intersect operation will be separated into two alternatives if the intersect range does not include any edges of the range (i.e. the intersect range cuts inside of the range). The next step is to reverse the final impact score to a value in the REMBRANDT level. Because a final impact score is mainly calculated from the geometric mean of alternatives, we reverse it by using the transformation function of the final score (f) as shown in (2).

$$r(f) = \frac{ln(f)}{ln(2)} \tag{2}$$

The last step is to create the merged comparison table. Each member M_{ij} of the comparison table (the comparative value of alternative i and j) is the difference of values in REMBRANDT level between alternatives i and j. One alternative table is mapped onto one of the regrouped sets of criteria according to its previous parent criterion. The merged comparison table of criterion and all corresponding alternative tables is called a *merged request*. Instead of multiple requests, we continue creating a requirement in the third step of the main process with the merged request.

3.5 Illustrative example

To illustrate the concept of user request in terms of criteria and alternatives, we give an example of user requests in the smart building system. We pick up four applications as follows: HVAC,

Table 1 Applications Comparison Table

Application	HVAC	Air quality	Switching	Security
HVAC	1	1	3	1/3
Air quality	1	1	3	1/3
Switching	1/3	1/3	1	1/5
Security	3	3	5	1

Table 2 HVAC request

Criteria	Thematic	Accuracy	Timeliness	Others
Thematic	1	5	3	9
Accuracy	1/5	1	5	9
Timeliness	1/3	3	1	5
Others	1/9	1/3	1/5	1

Thematic	Set 1	Others
Set1={Temperature&heat, Movement&occupancy}	1	9
Others	1/9	1

Accuracy	High	Acceptable	Low
High [0.8,1]	1	3	5
Acceptable [0.6,0.8)	1/3	1	3
Low [0,0.6)	1/5	1/3	1

Timeliness	High	Meduim	Low
High ratio [0.7,—)	1	3	7
Medium ratio [0.3,0.7)	1/3	1	5
Low ratio [0,0.3)	1/7	1/5	1

Table 3 Air-quality-and-window-control request

Criteria	Thematic	Spatiality	Accuracy	Timeliness	Others
Thematic	1	1	5	3	9
Spatiality	1	1	5	3	9
Accuracy	1/5	1/5	1	1/3	3
Timeliness	1/3	1/3	3	1	5
Others	1/9	1/9	1/3	1/5	1

Thematic	Set 1	Others
Set1={Movement&occupancy, Smoke&gas, Status}	1	9
Others	1/9	1

Spatiality	Set 1	Set 2	Others
Set1={41-50}	1	3	5
Set2={81-100,121-130}	1/3	1	3
Others	1/5	1/3	1

Accuracy	High	Acceptable	Low
High [0.8,1]	1	3	5
Acceptable [0.6,0.8)	1/3	1	3
Low [0,0.6)	1/5	1/3	1

Timeliness	High	Meduim	Low
High ratio [0.7,—)	1	3	5
Medium ratio [0.3,0.7)	1/3	1	3
Low ratio [0,0.3)	1/5	1/3	1

Table 4 Switching-device request

Criteria	Thematic	Spatiality	Accuracy	Others
Thematic	1	1	5	9
Spatiality	1	1	5	9
Accuracy	1/5	1/5	1	3
Others	1/9	1/9	1/3	1

Thematic	Set 1	Others
Set1={Movement&occupancy, Status}	1	9
Others	1/9	1

Spatiality	Set 1	Set 2	Others
Set1={45-60}	1	3	5
Set2={131-140}	1/3	1	3
Others	1/5	1/3	1

Accuracy	High	Acceptable	Low
High [0.8,1]	1	3	5
Acceptable [0.6,0.8)	1/3	1	3
Low [0,0.6)	1/5	1/3	1

Table 5 Security-and-safety request

Criteria	Thematic	Spatiality	Accuracy	Others
Thematic	1	1	3	9
Spatiality	1	1	3	9
Accuracy	1/3	1/3	1	7
Others	1/9	1/9	1/7	1

Thematic	Set 1	Others
Set 1={Temperature&heat, Movement&occupancy, Smoke&gas, Status, Glass break}	1	9
Others	1/9	1

Spatiality	Set 1	Others
Set 1={1-20, 151-160, 181-200}	1	9
Others	1/9	1

Accuracy	Extreme	High	Acceptable	Low
Extremely High [1,0.9]	1	3	5	9
High [0.8,0.9)	1/3	1	3	7
Acceptable [0.6,0.8)	1/5	1/3	1	5
Low [0,0.6)	1/9	1/7	1/5	1

air quality and window control, switching devices, and security and safety. Each application defines a different kind of request and holds a different priority. Table 1 shows an application comparison matrix where the priorities of applications are specified.

We define criteria by using the VoI attribute taxonomy in [2]. Without any impact to our study, we assume that all sensor data are from the trustworthy sensors and already in the compatible format. We use four criteria over the content of data as follows: thematic relevance, spatiotemporal relevance, accuracy, and timeliness. We determine the thematic-relevance criterion by using the sensor types according to cross-tabulated smart building applications and sensors in [16]. For example, in HVAC application, data from temperature-and-heat detectors and movement-and-occupancy sensors will have thematic-relevance values higher than the others. For simplicity of the spatiotemporal criterion, we assume that all sensors are statically settled. So, this criterion will leave just spatial dimension. Table 2 shows HVAC application request. The request consists of a comparison matrix of criteria, followed by those of alternatives. HVAC is an application for providing indoor thermal comfort and acceptable air quality. It normally concerns about three criteria as shown in criteria table: Thematic, Accuracy, and Timeliness. According to the Saaty's scale, the criteria table of HVAC request shows that it give the priority of the thematic criterion more than that of timeliness and that of accuracy in one step (i.e. somewhat more important) and two steps (i.e. much more important) respectively. As seen, for each criterion, their attributes are defined. For example, the Thematic criterion consists of two alternatives, "Set 1", which represents the considered types of sensors including *temperature and heat* and *Movement and occupancy*, and "Others" (i.e. the other types of sensors). The accuracy criterion has three alternatives: "High", "Acceptable", and "Low". In particular, for the accuracy criterion, alternatives are represented in the range of a floating number valued from 0 to 1 (i.e. type-range alternative). For the timeliness criterion, it is the most complex one. Different from the first three criteria, which is statically measured from the properties of data, this criterion dynamically changes over the time. We define the timeliness value as a ratio of the number of data that is selected and sent at the gateway to the number of generating data over a specific period of time. However, the delay of sending and receiving at the gateway is omitted.

From the definition mentioned above, we create the input requests, including application comparison table, HVAC request, Air-quality-and-window-control request, switching-device request, and security-and-safety request. The last three application requests are shown in Table [3,4,5] respectively.

Table 6 Merged requirement

Criteria	Accuracy	Spatiotemp	Thematic	Timeliness
Accuracy	1	0.36	0.23	3.15
Spatiotemporal	2.81	1	0.39	4.97
Thematic	4.37	2.56	1	6.53
Timeliness	0.32	0.20	0.15	1

Thematic						
Glass break	1	0.18	0.36	0.27	0.36	4.58
Movement and occupancy	5.46	1	3.68	2.79	3.68	9.05
Smoke and gas	2.79	0.27	1	0.53	1	6.37
Status	3.68	0.36	1.89	1	1.89	7.26
Temperature and heat	2.79	0.27	1	0.53	1	6.37
Others	0.22	0.11	0.16	0.14	0.16	1

Accuracy				
R: [0.0, 0.6)	1	0.26	0.17	0.15
R: [0.6, 0.8)	3.91	1	0.33	0.26
R: [0.8, 0.9)	5.92	3.01	1	0.53
R: [0.9, 1.0]	6.81	3.90	1.89	1

Timeliness			
R: [0.0 0.3)	1	0.25	0.17
R: [0.3 0.7)	4.02	1	0.33
R: [0.7 —]	6.03	3.01	1

Spatiotemporal							
{131-140}	1	0.33	0.54	0.41	0.78	0.78	1.28
{1-20, 151-160, 181-200}	3.00	1	2.14	1.57	2.72	2.72	3.29
{41-44}	1.86	0.47	1	0.64	1.57	1.57	2.14
{45-50}	2.43	0.64	1.57	1	2.14	2.14	2.72
{51-60}	1.29	0.37	0.64	0.47	1	1	1.57
{81-100, 121-130}	1.29	0.37	0.64	0.47	1	1	1.57
Others	0.78	0.30	0.47	0.37	0.64	0.64	1

4. Experimental Results

We have used a network simulator Scenargie version 1.8 for simulating our scenario. This is a commercial simulator. Its powerful GUI and precise modeling of protocol sets are quite beneficial to a variety of simulator users. We have implemented the gateway-side computation (the value aggregation, sorting and selection algorithms) in C++ and incorporated them into the simulator code. However, the server-side computation (the merging algorithm) is implemented in Java considering future usage in real systems.

4.1 Simulation setup

To evaluate our system, we have applied our method to the smart building system described in Section 3.5 and set up the significant inputs of the simulation including sensor data, application requirements, resource constraints as follows.

4.1.1 Sensor Data

For the sensor data, the descriptions that are related to value assessment consist of sensor IDs, types, sampling rates, and accuracy. We generate data from five types of 200 sensors with the same sampling rate ($100ms^{-1}$). Each type has the same number of sensors (*i.e.* 40 sensors for each type). The sensor IDs are running numbers from 1 to 200 following this type order; (i) temperature and heat, (ii) movement and occupancy, (iii) smoke and gas, (iv) status and (v) glass-break. However, the sequences of data are randomly shuffled. We randomly generated accuracy for each sensor.

4.1.2 Application Requirements

A merged requirement consists of definitions for classifying regrouped alternatives in each mapped criterion and a reference score table of the merged request. An individual application request is defined in Section 3.5. After applying the merging algorithm as described in Section 3, we obtain a merged requirement and a score reference table as shown in Table 6 and Table 7, respectively. We use this merged requirement as an input at gateway for making a selection decision in simulation. Except for the timeliness criterion where the values for selection decision should lead the current low values to higher values, we define a function f of the current value x to a special value for the selection process by using the highest impact alternative (HIA) and the lowest impact alternative (LIA) as shown in (3).

Table 7 Score-reference table of merged requirement

Criteria	Alternatives	Score
Thematic	Movement and occupancy	3.043
	Status	1.658
	Smoke and gas	1.225
	Temperature and heat	1.225
	Glass break	0.668
	Others	0.198
Spatiality	{1-20, 151-160, 181-200}	1.306
	{45-50}	1.166
	{41-44}	1.041
	{81-100, 121-130}	0.930
	{51-60}	0.930
	{131-140}	0.879
	Others	0.830
Accuracy	R: [0.9,1]	1.288
	R: [0.8,09)	1.172
	R: [0.6,0.8)	0.949
	R: [0,0.6)	0.698
Timeliness	R: [0.7,—)	1.125
	R: [0.3,0.7)	1.017
	R: [0,0.3)	0.874

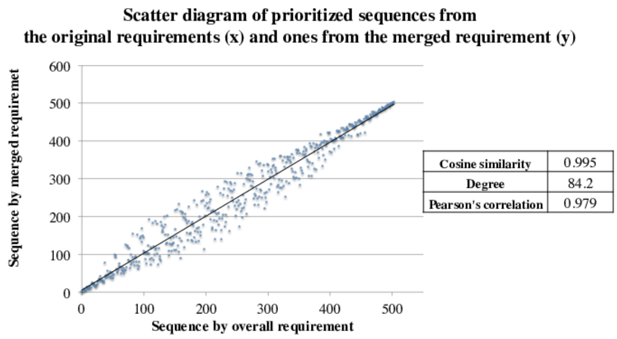

Fig. 5 Similarity and correlation of merging algorithm

$$f(x) = score(LIA) + \frac{score(HIA) - score(x)}{score(x)} \quad (3)$$

4.1.3 Resource constraints

We define the buffer size in term of the number of sensor data. We varied the available bandwidth related to the amount of incoming data per second from 200 sensors. To send all incoming data, the gateway requires a bandwidth at least 512 kbps in this case study. Therefore, we used the variation of bandwidth as follows: 128, 256, 384, and 512 kbps with unlimited buffer size, and determine the buffer size as the values starting from 100 to 800 with unlimited bandwidth.

4.2 Simulation Results

Our results are all in term of satisfaction values. The satisfaction value is defined as a percentage of the final score over the highest possible value calculated from a requirement. It will be 100% when the sensor data contain values which match to the highest-score alternative for all criteria. Sending all data does not mean the satisfaction value will become one which reflect the fact that not all generated data are needed. We compared our method to the traditional First-In-First-Out method (*Non-Criteria*: NC) and the method that uses only the-highest-impact criteria (*Top-Criterion*: TC). According to the similarity and correlation result shown in Fig. 5, which uses a set of all possible combinations of alternatives in all criteria as an input instead of a randomly generated one, a merged request can provide a prioritized sequence that has 0.995 (84.2°) of cosine similarity and 0.979 of Pearson's correlation to one that is derived from the original application requests with priority consideration. So, we can use the merged requirement for representing overall original requirements.

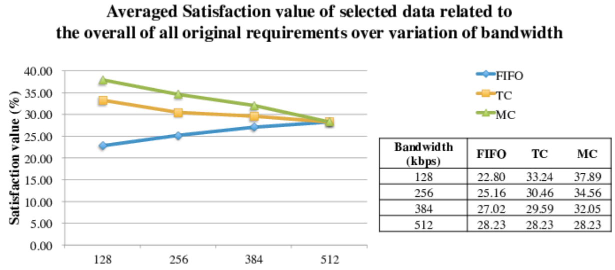

Fig. 6 Satisfaction value over variation of bandwidth

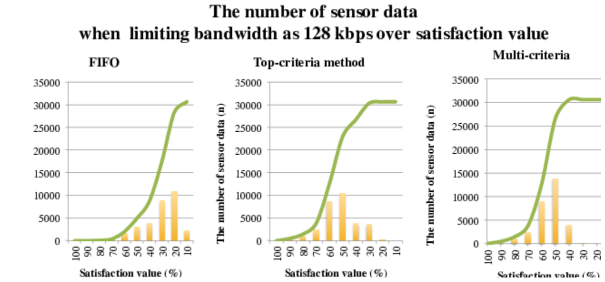

Fig. 7 Accumulated graph over satisfaction values when bandwidth is low

For the variation of bandwidth result, see Fig. 6. Without consideration of buffer size, we can see that our *Multi-Criteria* (MC) selection method results in significantly a higher satisfaction value compared to the others in the low bandwidth situation. All averages of satisfaction values using TC reside between that of using NC and MC. However, when bandwidth becomes larger, all methods converge to the same value at the situation that the bandwidth is sufficient for sending all data (512 kbps).

When looking at a low-bandwidth condition (128kbps), Fig. 7 shows the number of sensor data in each range of satisfaction value with an accumulated line. The growth of an accumulated line for the non-criteria method is low at a high value and high at a low value while the accumulated line of multi-criteria method has a high growth in a high value until moderate and then keeps flat in low. The growth trend of the top-criteria method is similar to that of the multi-criteria method, but, it is a little bit slower at the middle value.

See more detail in the result of the low bandwidth scenario, Table 8 shows the results of first three data which was sent to the server for each method. Note that the Data ID (DID) refers to the order of the data assigned by the counter at the gateway and SID is an abbreviation of Sensor ID. Obviously seen from the value of these first coming data, Multi-criteria method can select the data with the value higher than both top-criterion method and non-criteria method especially the latter one. Numerically, from the average value of the first three data, the value of data from multi-criterion method is higher than the top-criterion data

Table 8 First three data which sent to the server when limiting bandwidth at 128 kbps for each method

Multi-criteria method

DID	SID	Sensor Type	Accuracy	Time	Value
44	45	Movement&occupancy	0.8	0.1	63.16
107	50	Movement&occupancy	0.8	0.1	63.16
79	43	Movement&occupancy	0.8	0.1	56.42
Average value = 60.91					

Top-criterion method

DID	SID	Sensor Type	Accuracy	Time	Value
199	76	Movement&occupancy	0.7	0.1	36.41
159	57	Movement&occupancy	0.8	0.1	50.36
158	54	Movement&occupancy	0.6	0.1	40.76
Average value = 42.51					

Non-criteria method

DID	SID	Sensor Type	Accuracy	Time	Value
0	175	Glass break	0.7	0.1	7.99
1	148	Status	0.8	0.1	24.52
2	141	Status	0.9	0.1	26.94
Average value = 19.82					

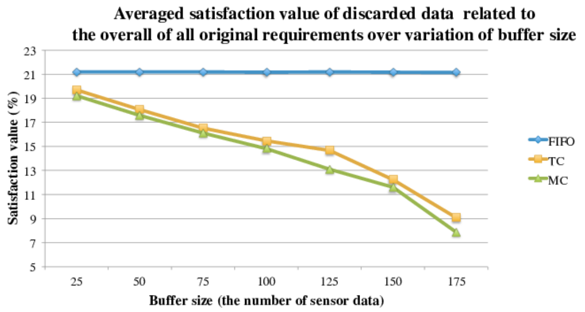

Fig. 8 Discarded satisfaction values over variation of buffer size

up to 43.29% and about 3 times of that of the data from non-criterion method. It may be noticed that all of the first three data of both Multi-criteria method and Top-criterion method contain only one sensor type, Movement&occupancy, while non-criteria method provide many types of sensor data. This is caused by the score of this type of sensor is much higher than the others about 2 times. However, different from the top-criterion, the multi-criteria method still gives a chance for the data from the sensors with other types to be sent to the server if they have higher score for other criteria even in this low bandwidth situation.

Then, we provide a sufficient bandwidth and evaluate the discarded values with a variation of buffer size. As shown in Fig. 8, the averages of discarded satisfaction values using multi-criteria are always less than that of others. In the same trend of the bandwidth variation, the results of using top-criterion method always reside between those of multi-criteria and non-criteria methods. In addition, for having criteria-concerned, we observe that the larger size the buffer has, the lower average value it will be lost. For more detail when the buffer size is low, the bar graph in Fig. 9 shows the comparison between the gained value and the dropped value. The worst method, non-criteria, causes the gained value nearly same as the dropped value. For the multi-criteria and

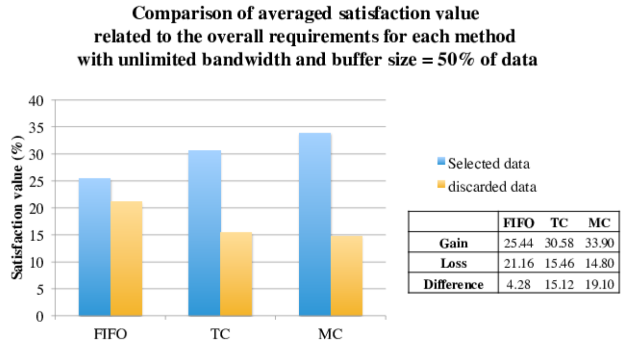

Fig. 9 Comparison of gained and dropped value when buffer is small

top-criterion methods, both of them have high gained-values and low dropped-values. Nevertheless, we do not recommend the top-criterion method even it seems to result in high efficiency. Obviously seen in the case where the top criterion is not time-dependent criterion, some data are possible to be in a starvation state. With the simulation given in this paper where the top-criterion is thematic, the selected data on the low bandwidth condition mostly contain one type, movement-and-occupancy, which is the highest-impact alternative. Moreover, according to the numerical value, we have also found that using multi-criteria method always gain better value and loss lower value compared to the results from the top-criterion method.

5. Discussion

There are many researches and theories about recommendation systems. However, only a few studies that applied them for sending process in WSNs. It is still open to study more about utilizing those well-known techniques. For improvement of user satisfaction, instead of assessing the value of each data like us, another option is assessing the value of each sensor. This option may reduce the complexities of data selection and bandwidth management. However, assessing the overall value in sensor level may loss some benefit from value inside of the sensed value. So, tradeoff between the importance of satisfaction and complexity should be concerned.

Our system can be further applied with other optimization techniques including data compression, event-driven system. Those techniques mostly consider only one requirement from one user while our system concerns more about multi-user scenario. Data compression techniques can be applied before sending selected data while event-driven techniques can be applied when collecting sensor data before selection. There can be more than one event from more than one user. With our proposed method, the results matched to those specific event will be prioritized.

Even though users do not have to think about the absolute value for the attributes for each criterion, they still have to define the comparison values between each attribute and each criterion. This may be a limitation for implementing our system in the environment that users are not allowed to define the requirement directly to the system (e.g. in shared sensor network system). We are now considering additional feature that can generate a requirement, compatible to our system, automatically. However it still allows users to send some feedbacks, in direct or indirect way, for adaption during operating period.

In addition, from the last scatter diagram, our merging algorithm still has a gap between the plot points and a trend line in the middle sequence. In the future work, we plan to find out more appropriate transformation function which still be less complex but more precise.

6. Conclusion

In this paper, we have introduced an approach to improve efficiency in term of application satisfaction in sending the large amounts of sensor data over a limited link. As an additional benefit to the well-researched compression techniques, we have proposed a data selection system based on the values of data for increasing satisfaction of multiple applications that utilize the sensor data. Application users can specify the desired data and then the data that meet user demand will be selected and sent back to the user faster. Moreover, in the case that availability of resources is so low that some data must be discarded, using our method can prevent the desired data in high priority. We have introduced a MCDA technique called REMBRANDT and evaluated the satisfaction values of our method compared to two naive methods. The results show that our method significantly obtains higher satisfaction values and discards lower values. In addition, to support multi-user system, we proposed merging process for merging multiple requests of users into one. We have evaluated the merging algorithm by comparing the prioritized sequences generated from the original requests and from the merged request. They have a high cosine similarity and high Pearson's correlation. This fact has shown that our method can be applied to multi-application scenarios.

References

[1] Bagale, J., Shiyanbola, A., Moore, J. and Kheirkhahzadeh, A.: Towards a Real-Time Data Sharing System for Mobile Devices, *Next Generation Mobile Apps, Services and Technologies (NGMAST), 2014 Eighth International Conference on*, pp. 147–152 (online), DOI: 10.1109/NGMAST.2014.19 (2014).

[2] Bisdikian, C., Kaplan, L. M. and Srivastava, M. B.: On the quality and value of information in sensor networks, Vol. 9, No. 4, pp. 1–26 (online), DOI: 10.1145/2489253.2489265.

[3] Estrin, D., Govindan, R., Heidemann, J. and Kumar, S.: Next century challenges: Scalable coordination in sensor networks, *Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking*, ACM, pp. 263–270.

[4] Great Britain and Department for Communities and Local Government: *Multi-criteria analysis a manual.*, Communities and Local Government (2009).

[5] Hamed Maleki: Comparison of the REMBRANDT system with the Wang and Elhag approach: A practical example of the rank reversal problem, Vol. 6, No. 1 (online), DOI: 10.5897/AJBM11.060.

[6] Hunkeler, U., Truong, H. L. and Stanford-Clark, A.: MQTT-S-A publish/subscribe protocol for Wireless Sensor Networks, *Communication systems software and middleware and workshops, 2008. comsware 2008. 3rd international conference on*, IEEE, pp. 791–798.

[7] Lootsma, F.: *The REMBRANDT System for Multi-criteria Decision Analysis Via Pairwise Comparisons Or Direct Rating*, Reports of the Faculty of Technical Mathematics and Informatics, Delft University of Technology, Fac., Uni. (1992).

[8] Olson, D. L., Fliedner, G. and Currie, K.: Comparison of the REMBRANDT system with analytic hierarchy process, *European Journal of Operational Research*, Vol. 82, No. 3, pp. 522–539 (1995).

[9] Pragnya, K. R. and Chaitanya, J. K.: Wireless Sensor Network based Healthcare Monitoring System for Homely Elders, Vol. 6, No. 5, pp. 2078–2083.

[10] Puccinelli, D. and Haenggi, M.: Wireless sensor networks: applications and challenges of ubiquitous sensing, *Circuits and Systems Magazine, IEEE*, Vol. 5, No. 3, pp. 19–31 (2005).

[11] Suryadevara, N. K. and Mukhopadhyay, S. C.: Wireless sensor network based home monitoring system for wellness determination of elderly, Vol. 12, No. 6, pp. 1965–1972.

[12] Suthaharan, S.: Big Data Classification: Problems and Challenges in Network Intrusion Prediction with Machine Learning, *SIGMETRICS Perform. Eval. Rev.*, Vol. 41, No. 4, pp. 70–73 (online), DOI: 10.1145/2627534.2627557 (2014).

[13] Tychogiorgos, G. and Bisdikian, C.: Selecting Relevant Sensor Providers for Meeting "Your" Quality Information Needs, IEEE, pp. 200–205 (online), DOI: 10.1109/MDM.2011.40.

[14] Tychogiorgos, G. and Bisdikian, C.: A framework for managing the selection of spatiotemporally relevant information providers, *Integrated Network Management (IM 2013), 2013 IFIP/IEEE International Symposium on*, IEEE, pp. 403–410.

[15] Virone, G., Wood, A., Selavo, L., Cao, Q., Fang, L., Doan, T., He, Z. and Stankovic, J.: An advanced wireless sensor network for health monitoring, *Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare (D2H2)*, pp. 2–4.

[16] Weber, V.: Smart Sensor Networks: Technologies and Applications for Green Growth (2009).

[17] Yoi, K., Yamaguchi, H., Hiromori, A., Uchiyama, A., Higashino, T., Yanagiya, N., Nakatani, T., Tachibana, A. and Hasegawa, T.: Multi-dimensional Sensor Data Aggregator for Adaptive Network Management in M2M Communications, *Proc. of IFIP/IEEE Int. Symp. on Intergated Network Management (IM2015)*, pp. 1047–1052 (2015).

[18] Zhang, Y., Yu, R., Xie, S., Yao, W., Xiao, Y. and Guizani, M.: Home M2M networks: architectures, standards, and QoS improvement, Vol. 49, No. 4, pp. 44–52.

