A Web-Based Sensor Data Management System for Distributed Environmental Observation

Takahiro Torii, Yusuke Yokota, and Eiji Okubo

Abstract—Recently, environmental observation gains in importance as primary sensor network application for actual use. To provide a practical environmental observation service, an appropriate mechanism for access and control of distributed sensor data and sensor networks is required. We propose a new web-based sensor data management system that extends the existing typical Sensor Web architecture for distributed environmental observation. Our system connects the Internet and remote places for environmental observation with networks that have relatively narrow bandwidth, such as cell-phone lines or satellite connections. It manages distributed sensor data in remote places and controls traffics between the Internet and remote places properly to provide practical utility to users of applications. We optimize our system for gathering requested sensor data on narrow bandwidth network. It allows users to gather sensor data on demand. Therefore users can obtain necessary and sufficient sensor data. On the other hand, it may fail when additional traffic occurs because of bandwidth restrictions. Consequently we introduce a query management mechanism for prevention of communication faults. We confirm this system can provide necessary and sufficient sensor data for multiple users through the simulation that estimates an appropriate network bandwidth to transfer sensor data from remote places.

Index Terms—Sensor Web, sensor network, distributed data management, query processing, environmental observation

I. INTRODUCTION

In recent years, applications of monitoring environmental information for observation of environmental changes or disaster prevention are becoming more important. A sensor network [1] is a key technology for monitoring them in open fields. However there is a challenge for management of sensor data if sensor networks are widely distributed in various remote places.

In a so-called Sensor Web [2] system, it provides access to sensor data on the Web. These sensor data are collected from sensor divides connected to the Internet. Users can access sensor data via Web clients anywhere at any time. Connecting environmental observation places to the Sensor Web system may cause a problem related to network performance. In many cases, it is difficult to prepare a stable and large-bandwidth network between the Internet and remote observation places. Although a typical Sensor Web system

collects all sensor data from sensor networks and stores them into a central storage on the Internet, it is difficult to collect all sensor data from sensor networks in remote observation places via the unstable and narrow-bandwidth networks because of overcapacity of network and increasing sizes of sensor data including image data in recent years. Therefore extension of existing typical Sensor Web architecture suitable for such condition of environmental observation is required. We propose a web-based sensor data management system for distributed environmental observation, which connects the Internet and remote places for environmental observation with networks that have relatively narrow bandwidth, such as cell-phone lines or satellite connections. It manages distributed sensor data in remote places and controls traffics between the Internet and remote places properly to provide practical utility to users of applications.

The rest of the paper is organized as follows. In the next section, related work is mentioned and discussed. In Section III, We propose new sensor data management system architecture for distributed environmental observation. The design of the system is discussed in detail in Section IV. Section V shows the simulation results of our system behavior, and Section VI discusses issues in operation of the proposed system. Finally, Section VII concludes the paper.

II. RELATED WORK

Reference [3] is a web-based sensor network system for agriculture support. One of a main component of the system is Field Server. It has various kinds of sensors including image sensors. Image data provides useful and helpful information for agricultural users to observe crop conditions. The system also has a web server to provide monitoring and analyzing applications at a remote site via the Internet for help farmers. Different from our proposed system, this system only will be applicable to observation places with sufficient network infrastructure. The restriction may limit the range of application of the system.

Cosm (https://cosm.com) is a kind of web service to share sensor data and environmental data on the Internet among users. It helps to make new applications and web services using these data. For example, it can map sensor data on the world map and visualize these sensor data. Difference of this system and ours is a data management mechanism. Cosm uses a centralized agent database on the Web. In contrast, our system has distributed databases in remote observation places to reduce traffic through unstable and narrow-bandwidth networks.

SWE architecture [4] is designed for creation of web-based sensor network systems. It makes all sensors and repositories

T. Torii is with the Graduate School of Science and Engineering, Ritsumeikan University, Kusatsu, Shiga 525-8577 Japan (e-mail: ttorii@sol.cs.ritsumei.ac.jp).

Y. Yokota and E. Okubo are with the College of Information Science and Engineering, Ritsumeikan University, Kusatsu, Shiga 525-8577 Japan (e-mail: yyokota@cs.ritsumei.ac.jp, okubo@is.ritsumei.ac.jp).

of sensor data discoverable, accessible, and controllable via the Web. This architecture is very large scale due to the general purposes and complicated. In comparison with SWE architecture, the purpose of our proposed system is focused on realization of distributed environmental observation in remote places utilizing Sensor Web architecture. The architecture of our system is rather simple than the SWE one.

III. OVERVIEW AND PURPOSES OF THE OUR SENSOR DATA MANAGEMENT SYSTEM

A. Backgrounds

Typical web-based sensor network systems collect all sensor data from sensor networks and store them on a database server on the Web. It helps users' access arbitrary sensor data via the Web. The upper side of Fig. 1 shows the architecture of such systems.

However, when sensor network systems are deployed in remote places for environmental observation, enough network infrastructures are not expected in many cases. Therefore it may be difficult to prepare enough network bandwidth between the Internet and remote places and manage large traffic of sensor data.

On the other hand, most users of the system do not require entire raw sensor data. In general, they use sensor data to know changes of conditions, some anomalous data, aggregated values of sensor data, and so forth. These data will be generated via the process of computation of raw sensor data and have smaller sizes than the raw data.

For example, when an observer wants to know temperature data at nighttime, s/he does not need sensor data collected during daytime nor other types of sensor data. In this case, it does not need to send all types of sensor data all day to a database server on the Web. Transmitting selected sensor data from remote places will save traffics of network paths between the Internet and remote places.

To manage environmental observation tasks with a web-based data management system, consideration of unstable and narrow bandwidth network paths between the Internet and remote places will be indispensable. We propose the system that meets the requirement based on on-demand query processing issued by users. The lower part of Fig. 1 shows the architecture of the proposed system. In our system, all data from sensor networks are stored into micro storages in remote places initially. When users issue requests, only required sensor data will be transferred from micro storages in remote places to the Internet efficiently. Therefore the amount of the network traffics in our system will be less than that of the existing systems.

If the user requests need large amount of data transfer, network congestion will be occurred without precise control of issuing queries. Such cases will be discussed in Section VI.
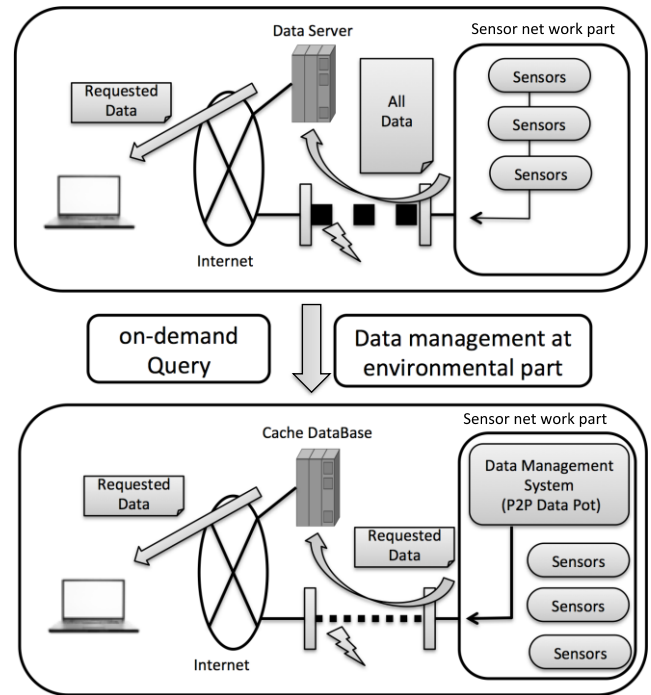

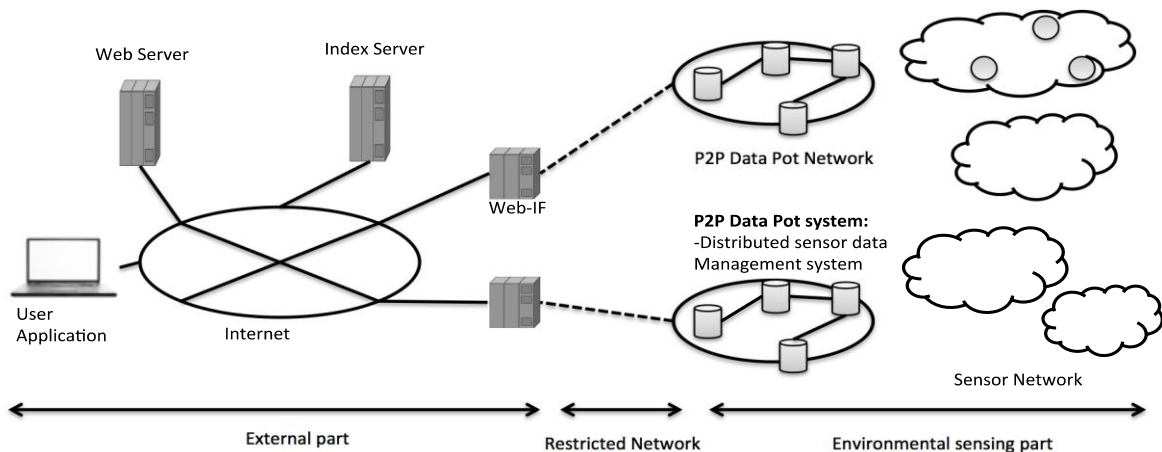Fig. 1. Architecture of web-based sensor network systems

Fig. 2. System overview.

B. System Overview

Fig. 2 illustrates the overview of the proposed system. It consists of Web Server, Index Server, several Web Interface hosts (Web-IFs), and several remote places for environmental observation/sensing. The roles of Web Server are accepting queries from user applications, retrieving corresponding sensor data from remote places, and sending them as results of the queries to user applications. Index Server manages meta data of sensor data and remote places for providing a look-up service to Web Server. It helps to hide internal information

about distributed sensing environment from users. Users can retrieve necessary sensor data without detailed knowledge of the remote sensor networks. The roles of Web-IF are an interface among Web and observation systems and to reduce redundant network traffics. It receives queries from Web Server, transfers them to a P2P Data Pot network at remote places, and returns results of the queries from the P2P Data Pot network to Web Server. It also has some optimization mechanism including data caching.

In remote places for environmental observation, P2P Data Pot networks and sensor networks are deployed. P2P Data Pot system [5] is consists of several small micro storage devices that have functionalities of wireless communication and data processing and organizes a P2P network on a wireless mobile ad hoc network for providing flexible system configurations. Fig. 3 shows the overview of the P2P Data Pot system. P2P Data Pot hides and abstracts distributed sensor networks and makes it easy to collect sensor data via Web. Each P2P Data pot is connected to a single sensor network and collects sensor data from there. Collected sensor data are stored into a storage of itself and shared among other Data Pots on a P2P network. P2P Data Pot has two roles in the proposed system. Firstly, it helps users to collect sensor data from various sensor networks. One P2P Data Pot network behaves as a single large database, thus users do not need consider internal network structures of remote places. Secondly, it realizes observation of extensive areas by bundling lots of sensor networks.

Web-IF on the Internet and a P2P Data Pot network in a remote place are usually connected via a narrow and unstable network, such as a cell-phone line.

C. Optimizing Data Transmission

To provide good responses to requests from users under the condition of limited network bandwidth, the proposed system optimizes data transmission. Web-IF has a caching mechanism for sensor data. In multi-user environment, requests from users will have many common parts of sensor data. Thus, Web-IF stores such parts of sensor data into its own cache. When cached data are requested, Web-IF returns the stored cache data without communication with remote places via a narrow bandwidth network and reduces unnecessary traffics.
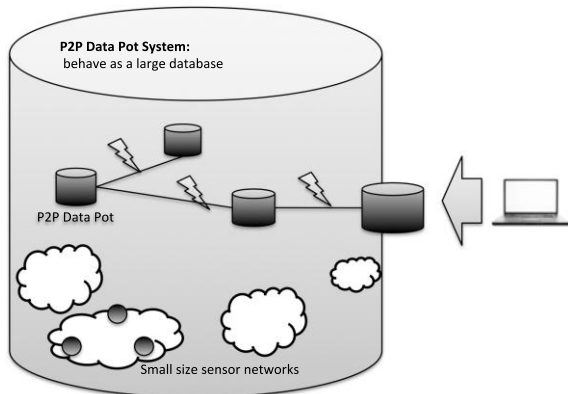

Fig. 3. P2P data pot system

D. Management of Distributed Sensor Data

In this system, collected sensor data are distributed to storages on P2P Data Pots in remote places. It is different from typical Sensor Web architecture that manages sensor data on a central database server. To manage such distributed sensor data efficiently, Index Server provides a look-up service to applications for searching necessary sensor data. It has metadata of all sensor data that include locations of sensors, types of sensors, and corresponding data access points (addresses of Web-IFs). This service hides the nature of complex distributed data from users and enables users to issue simple SQL-like queries. When such queries are issued from users, the system accesses the look-up service of Index Server, then decomposes and distributes the queries to appropriate Web-IFs. Each Web-IF collects requested data from a P2P Data Pot network in remote places or its own cache and returns them. Finally, separated query results are integrated and sent to the users.

IV. DESIGN

This system consists of five modules. The design of the module organization is focused on optimization of data transmission. In this section, functions of each module and a procedure of query processing are described.

A. Modules

Fig. 4 shows main modules of the system. It consists of Client, Web Server, Index Server, Web-IF, and GW-DP.
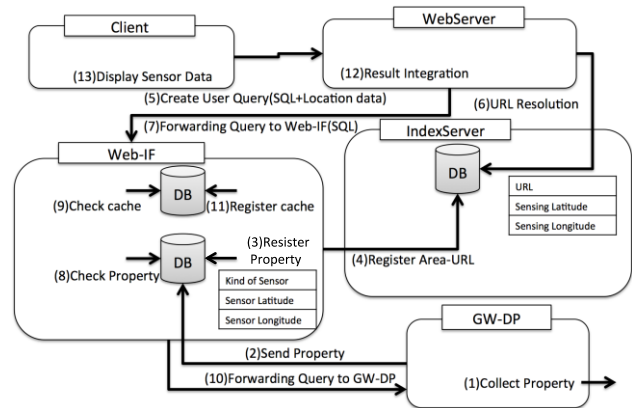

Fig. 4. Relationship of system modules

- Client

Users can issue queries via client applications. Client provides functions to generate and issue queries using GUI. Users can make SQL-like queries with specifying sensor types, time and location information. Generated queries will be sent to Web Server.

- Web Server

Web Server plays a role of a center of query processing in the system. When it receives a query from Client, it decomposes the query and sends the decomposed queries to appropriate Web-IFs, then receives query results from Web-IFs. Finally, it integrates the query results into a single query result and sends the result to Client. To communicate with Web-IFs, URL information of Web-IFs is required. Thus, Web Server sends requests to Index Server to obtain URL information lists of Web-IFs.

- Index Server

Index Server manages URL lists of Web-IFs and location

information database of sensor networks corresponding to Web-IFs. Index Server provides a look-up service of Web-IFs to Web Server. When Web Server sends location information of sensor data to Index Server, Index Server replies URL information of corresponding Web-IFs to Web Server.

- Web-IF

The role of Web-IF is an interface between the Web and remote environmental observation places. It transfers queries from Web Servers to remote places, receives query results from remote places and returns the results to Web Servers. Web-IF also has a caching mechanism to store frequently used sensor data into its own caching database. If there are suitable sensor data in the caching database corresponding to a query from Web Servers, Web-IF returns the cached data. It omits unnecessary communication with remote places and saves traffics. Our system supports multi-user accesses and queries from multiple users are supposed to have similarities. It expects to work the caching mechanism effectively.

Web-IF also manages metadata of corresponding sensor networks in remote places: types of sensors, location of sensor nodes, and so forth. Thus Web-IF can judge validity of queries received from Web Server based on the metadata. For example, a query that requests humidity data will be invalid if corresponding sensor networks only have temperature sensors. Web-IF eliminates invalid queries and omits unnecessary communication with remote places to save traffics. Valid queries will be transferred to GW-DP in remote places.

- GW-DP

GW-DP (GateWay Data Pot) is a representative for a P2P Data Pot network in a remote place. It communicates with Web-IF on the Web and sends sensor data collected from the P2P Data Pot network.

B. Query Processing

The following explanations describe query processing procedure in the system based on Fig. 4.

1) Collect Property: GW-DP collects property information as metadata of sensor networks from a P2P Data Pot network.

2) Send Property: GW-DP sends property and location information to Web-IF.

3) Register Property: registration of property information on the database of Web-IF.

4) Register Area-URL: Web-IF registers its URL information to the database of Index Server.

5) Create User Query: Create a SQL-like query specifying location and time of sensor data by a user.

6) URL Resolution: requesting a URL list to Index Server.

7) Forwarding Query: Web Server transfers queries to all corresponding Web-IFs.

8) Check Property: judging to accept or decline a query based on property information.

9) Check Cache: To reduce traffics in a narrow-bandwidth network, Web-IF searches the requested sensor data from its own cache database.

10) Forwarding Query to GW-DP: Web-IF transfers a query to GW-DP.

11) Register Cache: Web-IF registers the sensor data sent from GW-DP to the cache database.

12) Result Integration: collecting and integrating all of results sent from Web-IFs.

13) Display Sensor Data: A user browses the integrated results of the query.

V. EVALUATION

In this section, a typical Sensor Web system and the proposed system are compared based on simulation. The comparison shows an appropriate network bandwidth to transfer sensor data from remote places.

The simulation method is as follows: queries that retrieve random sizes of sensor data will be issued using a certain period of time. The changes of turnaround times of queries in response to changes of network bandwidth are observed.

Table I shows parameters of simulation environment. A typical Sensor Web system needs 4.0Mbps to retrieve sensor data in this environment.

TABLE I: SIMULATION SENSOR NETWORK ENVIRONMENT

Kind of sensor	Number of sensor	Data size	Sensing cycle
Temperature	1000	20 bytes	10[s]
Image	1000	100Kbytes	20[s]

TABLE II: PARAMETERS OF NETWORK AND QUERIES (1ST SIMULATION)

Band width	Query cycle	Query data size
2.5 Mbps	5 minutes	0.1~10Mbit

Table II shows parameters of a network and queries in the first simulation.

Fig. 5 shows sizes of issued queries during simulation. Fig. 6 shows turnaround time of issued queries during simulation.
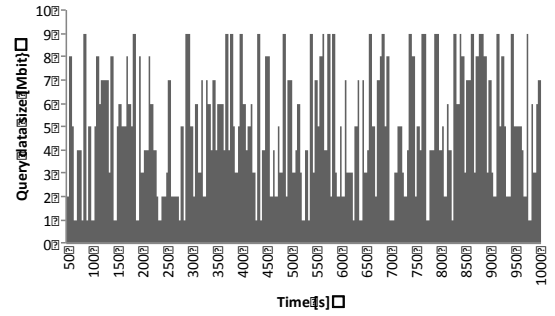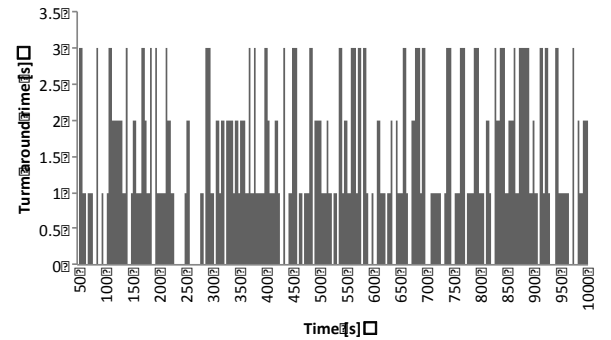

Fig. 5. Query data sizes (1st simulation)

Fig. 6. Query turnaround times (1st simulation)

In this case, all values of turnaround time are within 3 seconds because of enough network bandwidth.

Table III shows parameters of a network and queries in the second simulation. The bandwidth of the network is reduced to 1.0Mbps.

TABLE III: PARAMETERS OF NETWORK AND QUERIES (2ND SIMULATION)

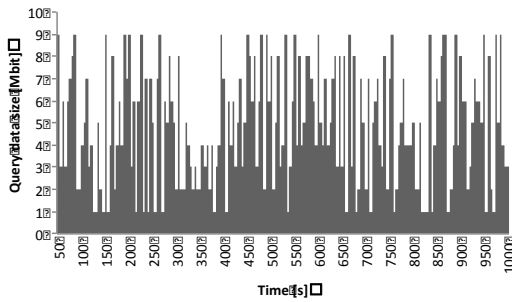| Band width	Query cycle	Query data size
1.0 Mbps	5 min	0.1~10Mbit

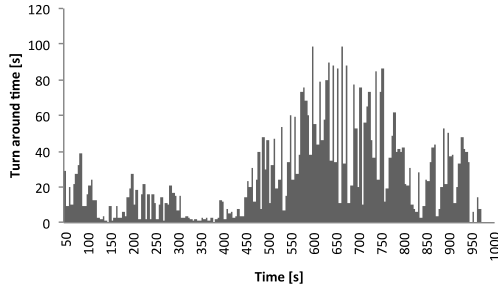Fig. 7. Query data sizes (2nd simulation)

Fig. 8. Query turnaround times (2nd simulation)

Fig. 7 shows sizes of issued queries during simulation. Fig. 8 shows turnaround time of issues queries during simulation. These results show the max turnaround time was 100 seconds. It might be happened by the concentration of queries that retrieves large size data.

This system processes all queries in parallel. Thus the network bandwidth is divided in a number of queries and delay time is extended in this case. As a result users are able to collect necessary sensor data by using the system in such environment that existing systems cannot provide the way to collect sensor data.

VI. DISCUSSIONS

A. Problem of Network Congestion

Our system is able to provide necessary sensor data requested from users or applications with narrower network bandwidth than a typical Sensor Web system on average. The on-demand query processing mechanism reduces total traffics of sensor data communication. However when very large sensor data are requested, the capacity of a network will reach the limit and some data will be dropped or long delay of data transfer may occur. The system should provide a mechanism to cope with such problems.

B. Delay Management

If applications or users issue queries and delay of response occurs, they cannot know when the results of the queries are arrived. It makes difficult to build applications dealing such exceptions properly.

In order to provide a notification service of turnaround time of queries to applications or users, a prediction mechanism that calculates approximate turnaround time of queries from specifications of them is required to implement on the system.

Applications can judge how to treat such queries with this information: canceling or rescheduling the query.

In order to calculate delay time, data size of the query in progress and requested data size of the query issued by an application are required. There are two methods to calculate turnaround time. First one is to calculate data sizes from specification of queries (number of sensors or amount of sensor data). Second one is to inquire data sizes to databases directly.

VII. CONCLUSION

In this paper, the extended Sensor Web architecture with unstable and narrow-bandwidth networks is proposed. Based on this architecture, an efficient sensor data management system for distributed environmental observation will be realized. The characteristics of this system are an on-demand query processing mechanism and a traffic optimization mechanism for unstable and narrow-bandwidth networks.

The simulation results show how much bandwidth is required to manage necessary and sufficient queries for applications. Discussions show requirement of a mechanism to predict approximate turnaround times of queries to cope with network congestion with a large amount of sensor data traffics.

REFERENCES

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, pp. 393–422, 2002.
[2] K. A. Delin, "The Sensor web: a macro-instrument for coordinated sensing," *Sensors*, vol. 2, pp. 270-285, 2002.
[3] O. Hukatsu, T. Kimura, and M. Hirafuji, "A Web-based sensor network system with distributed data processing approach via web application," *Computer Standards & Interfaces*, vol. 33, pp. 565-573, 2011.
[4] M. Botts, G. Percivall, C. Reed, and J. Davidson, "OGC sensor web enablement: overview and high level architecture," OGC Sensor Web Enablement, OGC 07-165, OpenGIS White Paper, 2007.
[5] Y. Fujisaki, K. Suzuki, Y. Yokota, and E. Okubo, "A cooperative storage system with wireless ad-hoc networking for wireless sensor networks," *IPSJ SIG Technical Reports*, 2008-MBL-44/2008UBI-17, pp. 149-156, March 2008.

Takahiro Torii received the B.E. degree from Ritsumeikan University in Computer Science in 2011. He is studying wireless sensor network systems at the postgraduate program of Ritsumeikan University.

Yusuke Yokota is an associate professor of College of University. He received his B.E. and M.E. degree in Computer Science from Kyoto University in 1996 and 1998, respectively, and a Ph.D. degree in Informatics from Kyoto University in 2005. His research interests include wireless sensor network systems, database technology, and CSCW. He is a member of IPSJ and ACM.

Eiji Okubo received his M.S. degree in Computer Science in 1977 from Hokkaido University. He received D.Eng. degree in Computer Science from Kyoto University in 1985. He is a professor of College of Information Science and Engineering of Ritsumeikan University. He has engaged in the study of operating system, database system, distributed system, and sensor network, among others. He is a member of IPSJ, ACM and IEEE-CS.