IWIN2010

International Workshop on Informatics

Proceedings of
International Workshop on Informatics

September 13-16, 2010
Edinburgh, Scotland, UK

Sponsored Informatics Society

IWIN2010

International Workshop on INformatics

Proceedings of

International Workshop on Informatics

September 13-16, 2010

Edinburgh, Scotland, UK

Sponsored by Informatics Society

URL: http://www.infsoc.org/

Publication office:
Informatics Laboratory
3-41, Tsujimachi, Kitaku, Nagoya 462-0032, Japan
Publisher:
Tadanori Mizuno, President of Informatics Society
ISBN:
978-4-902523-23-2
Printed in Japan

Table of Contents

Session 3: Network1 (Chairs Yoshia Saito and Tomoya Kitani) (Sep.14 9:00-12:30)

Invited Speech: Speaker Dr. Hiroshi Shimodaira (Chair Kouji Yoshida) (Sep.14 13:30-14:30)

Session 4: Network2 (Chair Yoh Shiraishi) (Sep.14 14:30-17:30)

Panel Discussion 1 (Chair Tomoya Kitani)
(Sep.15 10:00-12:00)

Panel Discussion 2 (Chair Tomoya Kitani)
(Sep.15 14:00-17:00)

A Message from the General Chair

It is my great pleasure to welcome all of you to Edinburgh, Scotland, UK, for the Forth International Workshop on Informatics (IWIN 2010). This workshop is sponsored by the Informatics Society.

The first workshop was held in Napoli, Italy, in September 2007, and the second workshop was held in Wien, Austria, in September 2008, and the third workshop was held in Hawaii, USA, in September 2009, where a variety of interesting papers were presented. In IWIN 2010, 29 papers were accepted and 18 papers were further selected as excellent papers which are considered having significant contributions in terms of the quality, significance, current interest among the professionals, and conference scope through the peer reviews by the program committees. Based on the papers, 4 technical sessions have been organized in a single track format, which highlight the latest results in research areas such as the mobile computing, networking, information system, and groupware & Education Systems. In addition, IWIN 2010 has an invited session from Dr. Hiroshi Shimodaira, who interested in statistical models for controlling talking faces and SSPNet. He is a lecturer of School of Informatics, and a member of Centre for Speech Technology Research (CSTR), University of Edinburgh.

I would like to thank all of participants and contributors who made the workshop possible. It was indeed an honor to work with a large group of professionals around the world for making the workshop a great success.

We are looking forward to seeing you all in the workshop. We hope you all will experience a great and enjoyable meeting in Edinburgh, Scotland, UK.

Kouji Yoshida

Kouji Yoshida
General Chair
The International Workshop on Informatics

Organizing Committee

General Chair

Kouji Yoshida (Shonan Institute of Technology, Japan)

Steering Committee

Toru Hasegawa (KDDI R&D Laboratories, Japan)
Teruo Higashino (Osaka University, Japan)
Tadanori Mizuno (Shizuoka University, Japan)
Jun Munemori (Wakayama University, Japan)
Yuko Murayama (Iwate Prefectural University, Japan)
Ken-ichi Okada (Keio University, Japan)
Norio Shiratori (Tohoku University, Japan)
Osamu Takahashi (Future University-Hakodate, Japan)

Program Committee

Kouji Yoshida (Shonan Institute of Technology, Japan)
Gen Kitagata (Tohoku University, Japan)
Hiroshi Mineno (Shizuoka University, Japan)
Hirozumi Yamaguchi (Osaka University, Japan)
Teruyuki Hasegawa (KDDI R&D Laboratories, Japan)
Tomoya Takenaka (Shizuoka University, Japan)
Teruhisa Ichikawa (Shizuoka University, Japan)
Yuko Murayama (Iwate Prefectural University, Japan)
Jun Sawamoto (Iwate Prefectural University, Japan)
Masahiko Ishino (Fukui University of Technology, Japan)
Yoshia Saito (Iwate Prefectural University, Japan)
Fumiaki Sato (Toho University, Japan)
Takaya Yuizono (Japan Advanced Institute of Science and Technology, Japan)
Taizo Miyaji (Tokai University, Japan)
Tomoya Kitani (Shizuoka University, Japan)
Hidekazu Shiozawa (Tamagawa University, Japan)
Tomoo Inoue (University of Tsukuba, Japan)
Hirosato Tsuji (Mitsubishi Electric Corporation, Japan)
Kozo Okano (Osaka University, Japan)
Toshihiro Shikama (Fukui University of Technology, Japan)
Shohei Yokoyama (Shizuoka University, Japan)
Hiroshi Shigeno (Keio University, Japan)
Yoh Shiraishi (Future University-Hakodate, Japan)
Shinichiro Mori (Fujitsu Laboratories, Japan)
Takuya Yoshihiro (Wakayama University, Japan)
Tsukasa Kudo (Shizuoka Institute of Science and Technology, Japan)
Masakatsu Nishigaki (Shizuoka University, Japan)
Satoru Fujii (Matsue College of Technology, Japan)
Yuji Wada (Tokyo Denki University, Japan)
Jun Munemori (Wakayama University, Japan)

Session 1: Information System (Chair Takaya Yuizono)

A Correction Reflected Query Method of Database during Online Entry

Tsukasa Kudo†, Yui Takeda‡, Masahiko Ishino*, Kenji Saotome**, Kazuo Mutou***, and Nobuhiro Kataoka****

†Faculty of Comprehensive Informatics, Shizuoka Institute of Science and Technology, Japan
‡Mitsubishi Electric Information Systems Corporation, Japan
* Department of Management Information Science, Fukui University of Technology, Japan
** Hosei Business School of Innovation Management, Japan
*** Faculty of Science and Technology, Shizuoka Institute of Science and Technology, Japan
**** Department of Information Technology, Tokai University, Japan
kudo@cs.sist.ac.jp

Abstract - The database of the mission-critical systems is updated with entry data by transaction processing, and are queried to make statistics and so on by batch processing generally. Such a batch processing had been executed at the overtime to avoid the data entry service time, because it occupied the database for hours. On the other hand, in recent years, the entry service time is being rapidly extended with the development of the Internet business. So, the methods to execute the both concurrently have been put to practical use. However, there are some cases that cannot be supported by only the conventional methods, because there are various kinds of database query and operation in the actual mission-critical system. In this paper, to support such the case, we propose a query method to query the database as of designated time reflecting the correction entered after the time. Moreover, we implemented this method into a mission-critical system, and confirmed the effect to reduce the overtime batch processing in the actual operation.

Keywords: temporal database, transaction time database, mission-critical system, query, integrity, batch processing

1 INTRODUCTION

In the mission-critical system such as the retail, the finance, the manufacture, because data are entered by many online terminals concurrently (hereinafter "online entry"), concurrency controls are executed by the transaction processing [5]. On the other hand, a great deal of data processing, such as periodic sum of entered data, is processed by the batch processing [5]. For example, in the retail system, sales information at stores is reflected into its database immediately by the transaction processing; on the other hand, the settlement of accounts is calculated by the batch processing. Here, the batch processing had been executed in night to avoid the time zone of the online entry, because it occupy the database for hours to process a great deal of data. However, in recent years, this time zone was expanded by the development of the internet business and so on. As a result, it often caused a problem that the batch processing didn't complete in the given time.

For this problem, the method to maintain the integrity of query result of database even during the online entry had been implemented. For example, the multiversion concurrency control of database [2], by which the integrity of query result is maintained during the online entry, is used widely. And, authors showed that the integrity of the snapshot of the bitemporal database can be maintained during the online entry in the actual mission-critical system, even in the case that error data were detected, by reflecting its correction into the query result [9]. Here, the bitemporal database is a kind of temporal database [7], [13], which manages both of the transaction time and the valid time. The former is the time that data is valid in the database; the latter is the time that data is valid in real world [4], [6], [11], [13]. And, its query target was the data at the designated valid time.

However, in the case of the settlement of accounts and so on, the processing target is the data that was online entered by the deadline time, which is the database status as of this time. And, if error data are detected, they have to be corrected while the processing. In this case, the multiversion concurrency control has the problem that does not support the reflection of data correction after the deadline time; the bitemporal database also has the problem not being suitable for such the system that the status of real world was not entered instantly.

Our goal in this paper is to provide the query method that maintains the integrity of query result with reflecting the data correction, even in the above-mentioned case. For this purpose, we propose the correction query method, which uses the transaction time. We show that the corrected data is queried without influences of the online entry by this method. Moreover, we implemented this method into an actual mission-critical system, and confirmed the effect to reduce the overtime batch processing.

The reminder of this paper is organized as follows. In section 2, we show the problem to intend for, and in section 3, we propose the query method to solve this problem. In section 4, we show an implementation case of this method in a mission-critical system, and in section 5, we evaluate the method based on the implementation result. Finally, we consider this method in section 6.

2 PROBLEM WITH BATCH PROCESSING

2.1 Constitution of Batch Processing

In the mission-critical system, a certain level integrity of online entered data is maintained by the integrity control of

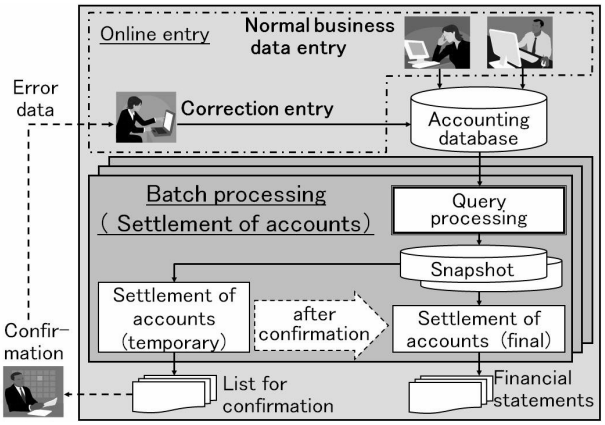


Figure 1: An example of batch processing constitution.

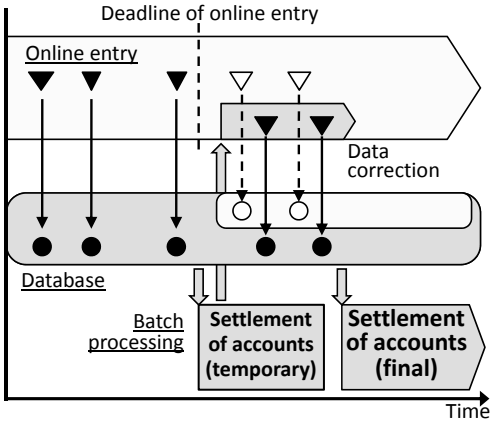


Figure 2: Settlement of accounts data by time series

the database management system and the transaction processing, and by the checking function of the business application program. In this paper, we define the integrity as what the state of the real world is reflected in the database with validity and completeness [10]. By the way, the integrity confirmation with querying a large quantity of data needs to be executed by batch processing. For example, the calculations of total for the collation with the actual cash or the actual articles, or the consistency check among some tables and so on. So, in the batch processing, the first process is usually the integrity confirmation of its target data.

Figure 1 shows the example of the batch processing about the accounting system. Accounting data is accumulated in the database by the online entry, and the settlement of accounts processing is executed regularly. In this processing, temporary processing is executed first to prevent errors of the processing, in which various kinds of data check is done. And, when error data is detected, it is corrected by the online entry. In this way, after all confirmation is complete, final processing is executed to make the financial statements.

Here, the query processing in the batch processing (hereinafter "batch query processing") of Figure 1 has to be executed without undergoing influence of the online entry, though it is executed concurrently with the online entry. So, even if

(1) Case of normal query

ID	Va	Vd	Ta	Td	Amount	Result
001	4/19	4/20	4/20	4/21	1,000	
001	4/19	4/20	4/21	now	1,500	●
002	4/20	4/21	4/21	now	3,000	

(2) Case of wrong query

ID	Va	Vd	Ta	Td	Amount	Result
001	4/19	4/20	4/20	4/21	1,000	
001	4/19	4/20	4/21	now	1,500	●
002	4/19	4/20	4/21	now	3,000	●

Figure 3: Query of correction data by bitemporal database

the correction data is entered by the online entry, it must be distinguished from the normal business entry data entered after the deadline time. Figure 2 shows the state of data of the settlement of accounts processing of Figure 1 by the time series. In figure 2, "▼" shows both of the online entry before the deadline time, and its correction entry; "●" shows the data corresponding to them. Also, "▽" shows the new online entry after the deadline time; "◯" shows the data corresponding to this. In the settlement of accounts processing, the temporary processing is executed for the confirmation about the data entered by the deadline time of Figure 2. And, the final settlement of accounts processing is executed after correction of the data error. Therefore, the target data of the settlement of accounts processing is the query result as of the the deadline time, in which only the correction entered after the time is reflected. That is, in Figure 2, only the data shown by "●" is the target for the final processing.

2.2 Problem about Conventional Database Query Method

We show the problem about the conventional database query method in the case of batch query processing accompanied by the data correction. In the multiversion concurrency control, the version of the database is managed with the time series. That is, the data entered after the deadline time for the correction cannot be distinguished from the normal business entry. Therefore, in the case shown in Figure 2, there is the problem that even the normal business entry data shown by "◯" become the processing target, too.

For this problem, we showed a solution utilizing the bitemporal database and confirmed that we could execute the batch processing even while the online entry in the actual mission-critical system [9]. In the bitemporal database, both histories of the valid time and the transaction time are managed, and the state of data, which once existed in the database, is accumulated as the records. That is, the both records of the state of the database and the real world are accumulated [4], [8]. For example, in the personnel management system of the company, the period that a person was in office for one duty position is shown with the valid time; on the other side, the period that its data was valid in the database is shown with the transaction time. Incidentally, the database that manages none of these times is called the snapshot database [12].

Figure 3 shows the application example of the bitemporal database to the travel expense checkout of the accounting system, in which correction data is queried on the condition that

the deadline time is April 20th. In Figure 3, "$[V_a, V_d)$" shows the period of the valid time, i.e. one business trip period, and "$[T_a, T_d)$" shows the period of the transaction time, i.e. the period that its slip data was valid in the database of the system. Incidentally, the time is expressed by the unit of a day. And, "●" of the column " Result" shows the queried data for the query condition explained below. On April 20th, the data $ID = 001$ was entered, and on April 21st, the correction entry of the data $ID = 001$ and the new entry of the data $ID = 002$ was done. Here, when making a travel expense checkout data aggregate $D = \{d\}$ and designating the valid time t_v and the transaction t_t, the following data is queried as the snapshot as of the above-mentioned time.

$$D_1 = \{d | d \in D, t_v \in [d[V_a], d[V_d]) \wedge t_t \in [d[T_a], d[T_d])\} \quad (1)$$

Here, $d[V_a]$ shows the instance of the attribute V_a in d, and the others are same, too.

Therefore, as shown in (1) of Figure 3, when time were designated as $t_v = April\ 19th$ and $t_t = April\ 21st$, the data $ID = 001$ after correction is queried; the data $ID = 002$ is not queried. Here, the time "now" of T_d shows the corresponding data is valid at the time to query [1], [14].

However, in the actual business, the state of the real world isn't always reflected into the database immediately. (2) of Figure 3 shows the case that the entry of the travel expense checkout has been late. Though the valid time period of the trip is $[4/19, 4/20)$, its data was entered on April 21th. In this case, there is a problem that the query result includes the data $ID = 002$, because it satisfies the condition of equation (1). But nevertheless it is the normal business entry data after the deadline time.

Moreover, there is the problem that some businesses don't need to manage the valid time. For example, the slips of the purchase and the payment of the accounting system are managed by the system, so their valid time as for the real world isn't managed usually. That is, the split table of the database doesn't need to take the composition of bitemporal database.

3 PROPOSAL OF QUERY METHOD TO REFLECT DATA CORRECTION

In this section, we propose the correction query to query the data as of the deadline time with reflecting only the correction entered after the time.

3.1 Correction Query

The correction query deals with the database that manages the transaction time, i.e. the transaction time database. The relation [3] of the transaction time database R is expressed as following.

$$R(K, T, A) \quad (2)$$

We show each attribute as follows.

- $K = \{K_1, ..., K_m\}$
  This expresses the set of attributes constituting the primary key of the snapshot queried by the designated transaction time.
- $T = \{T_a, T_d\}$
  This expresses the time period attribute of the transaction time, which is generated by system and isn't made public to the users. Here, T_a shows the time that the data was added to the database (hereinafter "addition time"), and T_d shows the time that the data was logically deleted from the database (hereinafter "deletion time"). As long as the data hasn't been deleted yet, the instance of attribute T_d is expressed by the above-mentioned "now".
- $A = \{A_1, ..., A_n\}$
  This expresses the other attributes.

We can query the snapshot at any designated transaction time, which is the state of the database at the time. When making the designated time t, the relation of this snapshot is expressed by the following equation.

$$Q(t) = \{q | q \in R \wedge q[T_a] \leq t \wedge t < q[T_d]\} \quad (3)$$

Here, $q[T_a]$ shows the instance of the attribute T_a of q, and $q[T_d]$ is similar, too. In the correction query, both of the snapshot at t_1 and t_2 are queried. Here, t_1 is the deadline time, and t_2 is the designated time after t_1. And, the correction query result is the data that reflected the corrections entered by the time t_2 into the snapshot of t_1. Hereinafter, we call t_1 "query time", t_2 "correction query time". Incidentally, it becomes $t_1 < t_2$ from above-mentioned presupposition.

The relation of the correction query for R is expressed by the union of the following S_1 and S_2, i.e. $S = S_1 \cup S_2$. Here, S_1 shows the data not being changed or deleted between t_1 and t_2. So, the correction query result is the same as the snapshot of t_1. The corresponding data is expressed by the following equation, because it exists at the both of t_1 and t_2.

$$S_1 = \{s | s \in Q(t_1) \wedge s \in Q(t_2)\} \quad (4)$$

On the other hand, S_2 shows the data being changed or deleted between t_1 and t_2. So, the correction query result is the data after the change or delete. As for the change, it is expressed by the following equation, because the data before and after change is connected by the primary key attributes $r[K]$ and $s[K]$. And, by this definition, the data deleted by the time t_2 isn't the target of the correction query.

$$S_2 = \{s | s \notin Q(t_1) \wedge \in Q(t_2) \wedge \exists r \in Q(t_1); r[K] = s[K]\} \quad (5)$$

Incidentally, the data of the correction query result is the subset of the snapshot $Q(t_2)$, which is entered by the usual transaction, so the consistency of the data is maintained.

Figure 4 shows the example of the correction query, of which query time is the transaction time $t = t_1$ and correction query time is $t = t_2$. In the entered data $ID = 301, 302\ and\ 303$, $ID = 302$ was changed, $ID = 303$ was deleted, on the other hand $ID = 304$ was added newly after the time t_1. (2) of Figure 4 shows the correction query result for these data. First, the data $ID = 301$ is queried based on the equation (4); $ID = 302$ after correction is queried based on (5). Second, $ID = 303$ that was deleted and $ID = 304$ that was newly added don't become the target.

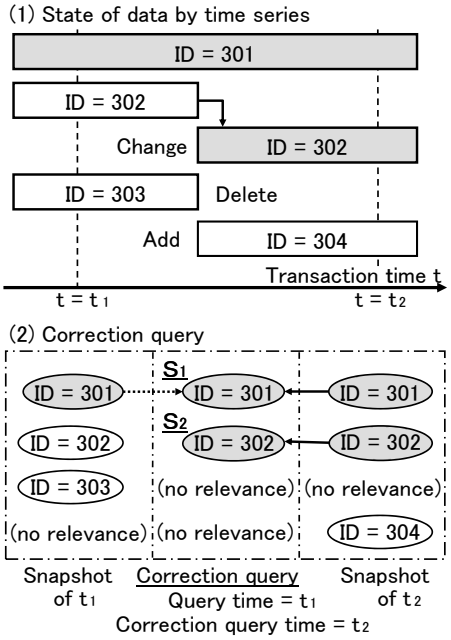


Figure 4: An example of correction query

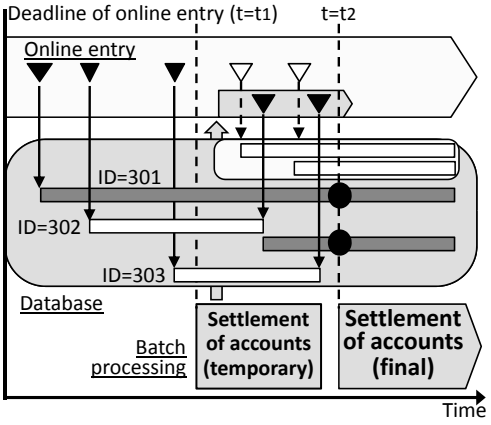


Figure 5: Correction query for settlement of accounts

3.2 Effect of Correction Query

We show that the problem shown in section 2.2 can be solved by the correction query. Figure 5 shows the application example of the correction query to the settlement of accounts processing, in the case of Figure 2. Here, we show the change of data of database by the time series like (1) of Figure 4. The temporary processing of the settlement of account had been executed for the data entered by the deadline time, and to correct the data, the change of $ID = 302$ and the deletion of $ID = 303$ were executed by the online entry based on the confirmation result of the temporary processing. On the one hand, the online entry of the normal business data were continued after the deadline time as same as before the time. In this example, the result of correction query, of which the query time is the deadline time $t = t_1$ and the correction query time is the start time of the "final" settlement of account processing $t = t_2$, is the data shown by " ● " in Figure 5. That is, the state of database as of the deadline time with reflecting the corrections entered after the time can be queried even during the normal business online entry without undergoing influence of this.

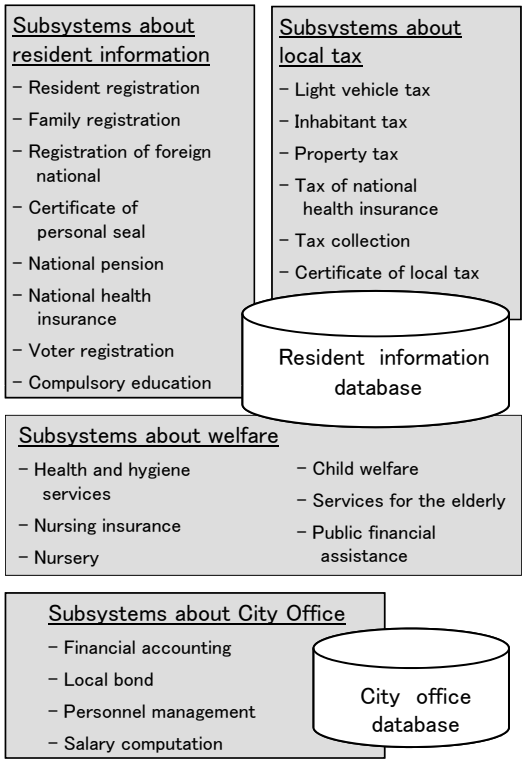


Figure 6: Composition of local government system

4 APPLICATION TO A MISSION-CRITICAL SYSTEM

In this section, we show the application result of the correction query to a mission-critical system, the local government system.

4.1 Overview of Local Government System

The local government system is a mission-critical system for the public administration business of the local government like a city hall. And, as shown in Figure 6, it consisted of various kinds of subsystems to assist the local government business. They were classified by business contents as follows.

(a) **Subsystems about Resident information**
They were used for the business, such as management and certificate of the residents who live in the city.

(b) **Subsystems about Local Tax**
They were used for the business of the local tax, such as levy and certificate about tax.

(c) **Subsystems about Welfare**
They were used for the business of welfare, such as qualification management, levy and grant.

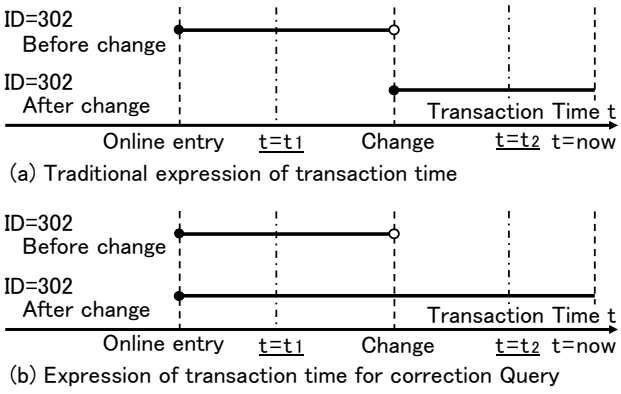


Figure 7: Implementation of transaction time

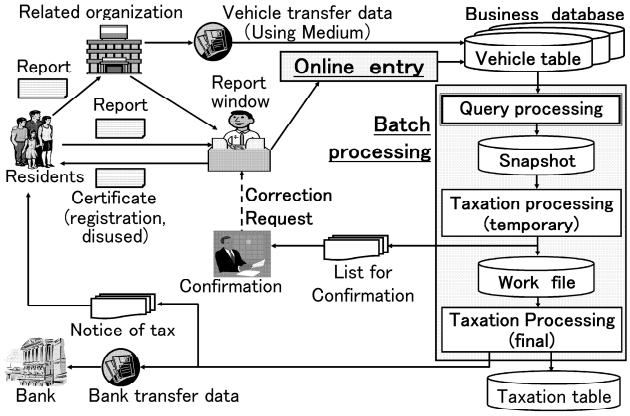


Figure 8: Dataflow of light vehicle tax business

(d) **Subsystems about City Office**
They were used for the business of the office work of local government, such as personnel management, salary computation and financial accounting.

In each subsystem, the reports were accepted at the report windows and online entered to accumulate in the database. And, the processing to query a large quantity of data was executed as the batch processing regularly or at any time. In the batch processing, the state of database as of the designated time was often queried. We show the example of batch processing like this below.

- **Population statistics:** based on the resident transfer reports, the statistics of such as the population and the number of households was made as of the end time of the first day of every month.
- **Taxation processing:** based on the reports about the local tax, the taxation processing was executed. It used the state of database as of the individually designated time.
- **Settlement of accounts processing:** based on the data of the income and the outlay, settlement of accounts processing was executed with the state of database as of the end time of every day, month and year.

4.2 Implementation of Correction Query

As shown in section 3.2, the correction query intends to the transaction time database. We used the commercial relational database and added the attributes of the addition time and the deletion time to each table to compose a transaction time database, depending on the necessity of the target business. Here, since transaction time is used as one of primary key attributes of the database, the unit of the transaction time had to be decided based on the frequency of data entry. In this system, data were entered from the terminals, and the data entry took several seconds at least. So, we made the unit of the transaction time 1 second. Incidentally, we made the attribute of the transaction time the closed information in users including the records as for it, so users could query only the latest state at the query time.

As for the change records with the transaction time, the data after change was conventionally expressed in the form, of which addition time was the changed time as shown in (a) of Figure 7. In the implementing of a correction query, it was necessary to connect the data before and after correction. So, the query processing became complicated if the conventional expression was used. To solve this problem, we implemented the transaction time with the expression, in which the addition time of the data after change is the time that the data was added first, as shown in (b) of Figure 7. Incidentally, in this expression, the deletion time becomes the primary key attribute; though, in the conventional expression, the addition time is the primary key attribute.

4.3 Composition of Subsystem for Business

As the example of the business system, to which we applied the correction query, we show the light vehicle tax subsystem that is one of the subsystems about the local tax. The light vehicle is taxed on the light vehicles, which is owned by the residents as of April 1st that is the basic date. And, the taxation processing is executed based on the data reported by the residents.

Figure 8 shows the dataflow of the light vehicle tax business. The acquisition reports of the light vehicles should be reported within 15 days; the disused and transfer reports should be reported within 30 days. However, these reports are accepted in the related organizations such as the Light Motor Vehicle Inspection Organization, the Land Transport Bureau and the light vehicle stores in addition to the report windows of the local government. The data accepted at the related organizations were delivered to the local government with paper reports for online entry or with mediums for lump-sum entry. For such operation, it often takes time to reflect the transfer data of the real world into the vehicle table of the system. Therefore, the taxation processing was executed for the data entered by the deadline time, and thereafter, tax correction processing was executed monthly for the data newly entered by the corresponding deadline time.

Online entry at the report windows could not be suspended during business hours, because the light vehicles license plate issue certificates or the disuse report receipt certificates had to

(1) Data of vehicle table on 5/6

ID	Owner	Ta	Td	5/6
001	Keiji, T.	5/6	now	●
002	Jouto, J.	5/6	now	●
003	Haisha, S.	5/6	now	●

(2) Data of vehicle table on 5/7

ID	Owner	Ta	Td	5/6	5/7	S
001	Keiji, T.	5/6	now	●	●	●
002	Jouto, J.	5/6	5/7	●		
002	Jouto, J.	5/6	now		●	●
003	Haisha, S.	5/6	5/7	●		
004	Tuika, F.	5/7	now		●	

Figure 9: Query result of vehicle table with correction query

be published immediately reflecting the reported data. On the other hand, the taxation processing and the tax correction processing were executed by the batch processing to make the tax payment notices to the residents and the account transfer requests to the financial institutions. So, to prevent the taxation error, the checklist and the statistics documents for the confirmation were made by the temporary processing first. And, when the data error was detected, it was corrected by the online entry. After this confirmation and correction, the final processing was executed.

So, the target data for the final processing was the state of database as of the deadline time, in which only the corrections after the time were reflected. Figure 9 shows the taxation processing case, of which the deadline time was May 6th and the execution time was May 7th. We show the state of database as of May 6th in (1) of Figure 9, and the data entered by this time was the target for the processing. We show the state as of May 7th in (2) of Figure 9, in which the change of the vehicle $ID = 002$, deletion of $ID = 003$ and addition of $ID = 004$ were reflected. Here, the transaction time of $ID = 002$ was implemented with the expression shown in (b) of Figure 7. In Figure 9, "●" of column "$5/6$" shows the snapshot data of May 6th; column "$5/7$" shows the snapshot data of May 7th; column "S" shows the correction query result, of which the query time was May 6th and the correction query time was May 7th.

As shown in the column "S", the correction query result was the snapshot at May 6th, in which only the correction entered by May 7th were reflected. So, the addition data $ID = 004$ was not included.

4.4 Combination with Other Query Methods

In the actual mission-critical systems, it is necessary to query the database in a wide range of conditions. For example, the light vehicle tax was paid by the tax notice or the bank transfer. Here, as for the bank transfer, it was requested by the resident with its transfer period. So, the data for the bank transfer needed to have the valid time attribute, and we had to implement it as a table of bitemporal database. On the other hand, we queried the master table by the multiversion concurrency control, because it was the table of the snapshot database without managing the transactiontime. In this way, as the constitution of the table was different with the condition of the target business, it was necessary to combine various kinds of query results to make the final outputs such as

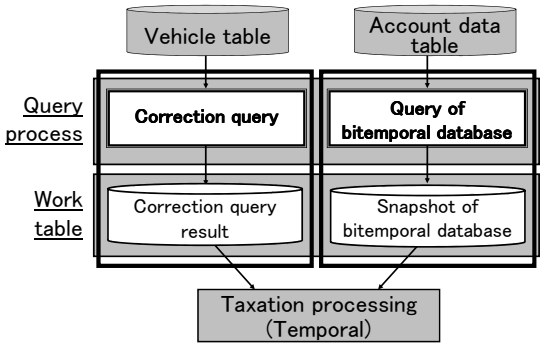


Figure 10: Combination with conventional query method

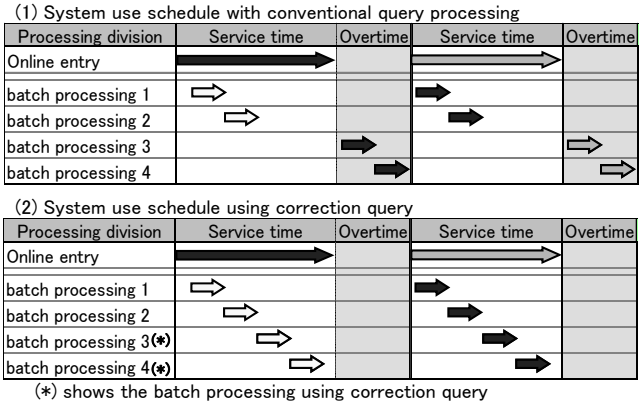


Figure 11: Reduction of overtime batch processing

the financial statements and so on.

In the application system, to solve this problem, we composed temporary files of the batch processing by the work tables, which are usually composed by the sequential access method (SAM) file. And, in the whole batch processing, we processed data by the query function of the database, to simplify each individual query procedure and maintain its performance. For example, as for the above-mentioned bank transfer, we queried the vehicle table by the correction query and queried the account data table by the snapshot of the bitemporal database on the other hand. Afterward, as shown in Figure 10, we combined these results by utilizing the query function of the database in the temporary processing executed next.

5 EVALUATION

5.1 Evaluation about Systems Operation

In the application system, online entry could not be suspended during business hours, because the certificates reflecting the entry data had to be published immediately as shown in Figure 8. On the other hand, conventionally, the batch processing using the data that took time until its entry or was not including the valid time data, could not be executed concurrently with the online entry. So, it had to be executed at the overtime like "batch processing 3" or "4" of (1) of Figure 11. In contrast, as the batch processing like this became able to

Table 1: Application rate of correction query.

No	business	total	T_a	T_d
(a)	resident	36	32(89%)	18(50%)
(b)	tax	72	58(81%)	31(43%)
(c)	welfare	40	37(93%)	24(60%)
(d)	office	63	42(67%)	8(13%)
	sum	211	169(80%)	81(38%)

be executed concurrently with the online entry by utilizing the correction query in the application system, it could be executed during the business hours on the next day as shown in (2) of figure 11.

As a result, all the batch processing to query database were executed during the business hours, and the overtime work could be reduced. Incidentally, the confirmation and the correction entry were also executed at the same time.

5.2 Evaluation about Coverage

Table 1 shows the application table number and rate of the correction query in the application system. We added the addition time T_a to the tables to manage the records with the transaction time; and we added the deletion time T_d to the tables for the correction in addition to T_a. Therefore, the rate of the column T_d of Table 1 is the application rate of the correction query. Here, the row number is the same as the subsystem classification number shown in section 4.1. And, it targets only the transaction table, so it excludes the following tables: the master tables such as the parameter table and the code table; the temporary data tables such as the work table; the derivation datas table such as the total sum.

Here, the table rate to have the addition time is 80%; the table rate to have the deletion time is 38%. That is, the correction query was applied to about 50% of the tables that manage the records. Here, the application rate depended on the subsystem. It was applied to only the 13% tables in the subsystems about the city office; on the other hand, it was applied to from the 43% to 60% tables in the other subsystems.

As shown in section 4.4, the queries with a wide range of conditions were necessary in the actual system operation. Table 2 shows the evaluation of the query method for these query condition. In addition, it shows the kind of the database corresponding to the query method, too. In table 2, "◯" shows that batch query processing can be executed during the online entry; "×" shows that there is the problem to execute the processing. The conventional query methods, i.e. the multiversion concurrency control and the snapshot of bitemporal database, have the problem for the query condition as of the designated transaction time with correction. By the correction query, we could execute the batch query processing even in the above-mentioned condition.

On the one hand, the multiversion concurrency control is necessary to query the tables of the snapshot database; the snapshot of the bitemporal database is necessary to query as of the designated valid time reflecting correction entry. Therefore, it is necessary to make the batch processing such a structure that can combine these query results for making the final output as shown in Figure 10.

5.3 Evaluation about Implementation

For the correction query was implemented in the query processing as shown in Figure 4, the online entry processing was same as before. And, as for the database table, we could implement the correction query easily, because we implemented the transaction time using the expression shown in (b) of Figure 7. For example, the correction query shown in Figure 9 could be executed by the following simple SQL.

$$select\ ID,\ Owner,\ T_a,\ T_d\ from\ Vehicle\ Table \\ where\ T_a \leq 5/6\ and\ T_d = now \quad (6)$$

In addition, there is the thing that plural history data are queried if T_d is designated as the past, not now. In this case, the history data that has earliest T_d becomes the query target. However, in the actual system operation, T_d was usually designated at "now", that is the time when the batch processing was executed. Therefore, such operation was unnecessary.

As shown in section 5.2, it is necessary to query the database in a wide range of conditions corresponding with the business needs and to combine these results to make the final output. For this problem, in the application system, we took the constitution of batch processing, in which we used the database work table instead of the SAM file as shown in Figure 10. As a result, we could combine them easier by using SQL function. By adopting the above-mentioned constitutions, in the application case to the local government of a population of about 40 thousand, the performance deterioration of query and online entry didn't occur comparing with the conventional method.

6 CONSIDERLATION

By the correction query, the problem of conventional query method, that is the query condition as of the designated transaction time with correction during online entry, could be solved. As the result of having applied it to an actual mission-critical system, we confirmed the effect that the overtime batch processing to query the database became unnecessary. In recent years, such the operation of mission-critical systems is increasing because of the rapid development of the internet business such as the electronic commerce, the electronic government and so on, in which users directly enter their data to the systems and the online entry cannot be suspended. So, the batch processing has to be executed in the online entry service time. Therefore, we consider that the correction query is effective, by which we can execute the batch query processing without suspending the online entry.

In the actual mission-critical systems, a wide range of data management and data query are necessary based on the business needs. So, it is necessary that the database can be queried by plural methods, and the final output has to be made by combining these query results. In particular, querying the database containing records is complicated. So, the method to meintain query performance is important. Therefore, we consider that our proposal method is effective: the implementation of the transaction time by the proposed expression; the

Table 2: Evaluation of query method with query time condition.

Target database	Query method	As of query start time	As of designated valid time with correction	As of designated transaction time with correction
Multiversion concurrency control	Snapshot database	○	×	×
Snashot	Bitemporal database	○	○	×
Correction Query	Transaction time database	○	×	◯

method using the work table to process the data step by step by utilizing the database function to simplify each query and combine their results to make final output.

The application rate of the correction query deeply depends on the subsystems as shown Table 1. Excepting the subsystems about the city office, because the subsystems deal the data based on the reports of real world, the wrong entry data has to be corrected as shown by the light vehicle business in section 4.3. On the other hand, as for the subsystems about the city office, the reports were often omitted in the business. For example, the slips of the financial accounting subsystem were managed in the database. So, when an approval slip was wrong, the new split was published for its adjustment. Therefore, we consider that the correction query is effective for the system that needs the internal correction to consistent its data with the state of the real world.

7 CONCLUSION

As for the system that takes time until the state of real world is reflected into its database, the batch processing is often executed using the data entered by the designated time. However, in this case, when the entry data is corrected, the integrity of the query result of the batch processing cannot be maintained during the online entry by the conventional query method. In this paper, we propose the correction query to query the data entered by the designated time with reflecting the corrections entered after the time. Moreover, we applied this to the mission-critical system and confirmed the effect to reduce the overtime batch processing in the actual systems operation.

Future study will focus on the development of the method, by which database can be updated with a large quantity of data in a lump during the online entry.

REFERENCES

[1] L. Bækgaard and L. Mark, Incremental Computation of Time-Varying Query Expressions, IEEE Trans. knowledge and Data Eng., Vol. 7, No. 4, pp. 583–590 (1995).

[2] P. A. Bernstein and N. Goodman, Multiversion Concurrency Control-Theory and Algorithms, ACM Trans. on Database Sys., Vol. 8, No. 4, pp. 465–483 (1983).

[3] E. F. Codd, Extending the database relational model to capture more meaning, ACM Trans. on Database Sys., Vol. 4, No. 4, pp. 397–434 (1979).

[4] N. Edelweiss, P. N. Hübler, M. M. Moro and G. Demartini, A Temporal Database Management System Implemented on top of a Conventional Database, Proc. XX International Conference of the Chilean Computer Science Society, pp. 58–67 (2000).

[5] J. Gray and A. Reuter, Transaction Processing: Concept and Techniques, Morgan Kaufmann, San Francisco (1992).

[6] C. S. Jensen, L. Mark and N. Roussopoulos, Incremental Implementation Model for Relational Database with Transaction Time, IEEE Trans. knowledge and Data Eng., Vol. 3, No. 4, pp. 461–473 (1991).

[7] C.S. Jensen, C.E. Dyreson and et. al., The Consensus Glossary of Temporal Database Concept – February 1998 Version, Temporal Database: Research and Practice.(the book grow out of a Dagstuhl Seminar, June 23–27, 1997), Lecture Notes in Computer Science 1399, Springer–Verlag, pp. 367–405 (1998).

[8] C. S. Jensen and R. T. Snodgass, Temporal Data Management, IEEE Trans. knowledge and Data Eng., Vol. 11, No. 1, pp. 36–44 (1999).

[9] T. Kudou, M. Ishino, K. Saotome, N. Kataoka and T. Mizuno, Implementation of Integrity Maintenance Method of Query Result by Bitemporal Database, International Journal of Infomatics Society, Vol. 1, No. 1, pp. 16–26 (2009).

[10] A. Motro, Integrity = validity + completeness, ACM Trans. on Database Sys., Vol. 14, No. 4, pp. 480–502 (1989).

[11] G. Özsoyoğlu and R. T. Snodgrass, Temporal and Real-Time Databases, A survey, IEEE Trans. knowledge and Data Eng., Vol. 7, No. 4, pp. 513–532 (1995).

[12] L. Shrira and H. Xu, SNAP: Efficient Snapshots for Back-in-Time Execution, Proc. 21st International Conference on Data Engineering., pp. 434–445 (2005).

[13] R. Snodgrass and I. Ahn, Temporal Databases, IEEE COMPUTER, Vol. 19, No. 9, pp. 35–42 (1986).

[14] B. Stantic, J. Thornton and A. Sattar, A Novel Approach to Model NOW in Temporal Databases, Proc. 10th International Symposium on Temporal Representation and Reasoning and Fourth International Conference on Temporal Logic, pp. 174–180 (2003).

One-click Peer-encouragement Mechanisms for Web-based Health Promotion System to Prevent Metabolic Syndrome

Takuya Yoshihiro†, Takuya Fujiwara‡, Etsuko Inoue†, Hiroko Tabe‡and Masaru Nakagawa†

†Faculty of Systems Engineering, Wakayama University, Japan
‡Graduate School of Systems Engineering, Wakayama University, Japan
†{tac, etsuko, nakagawa}@sys.wakayama-u.ac.jp

1 Introduction

Currently, metabolic syndrome has been generally regarded as one of the most serious diseases over the world. Although there are several definitions per country, metabolic syndrome is generally defined as the state that people put on significant weight [1]. Unfortunately, the prevalence of metabolic syndrome is considerably high all over the world [2]. The problem is that the people of metabolic syndrome have higher risk to be cardiovascular diseases or diabetes in the future, resulting in higher medical cost. Once people come to be cardiovascular disease or diabetes, their medical cost gets far higher. The importance to prevent metabolic syndrome is widely recognized.

Therefore, various activities to reduce or prevent metabolic syndrome has been tried all over the world. Especially in Japan, the government established the law to prevent metabolic syndrome under which many metabolic people are trying to reduce their weight under instruction of healthcare nurses. Recently in the instruction scene, several information systems are applied to improve the effect over reduction of users' weight. Typically in such systems, users report everyday's effort on reducing their weight through the Internet, and healthcare nurses periodically make proper advices according to their daily report in order to motivate further continuous effort of them.

This activity is actually effective to keep users' motivation for the everyday's effort to improve their weight. However, this style of instruction still requires much labor of health care nurses if they want to keep users' motivation at high level. One of the challenges in this area is to motivate users to continue their effort to reduce weight without increasing the labor of health care nurses.

In this paper, we propose the peer-encouragement mechanisms among users in which users encourage one another to improve motivation for everyday's effort by themselves without healthcare nurses. By this peer-encouraging mechanism, more users will keep their motivation to continue their activity without increasing healthcare nurses' labor. In our method, encouragement is done with simple "one-click" operation so that people can encourage others without stress and labor. In this paper we present the concept of peer-encouragement and the system design to be applied into practice.

The rest of this paper is organized as follows: In Sec. 2 we present the background of metabolic syndrome and the current activities to prevent/cure people against it, including the related information systems to help the activities. In Sec. 3 we present the concept and the mechanism of peer-encouragements and also its intended effects. In Sec. 4 we show the design of the system that we developed and explain how they work. In Sec. 5, the evaluation results for this system are presented, and we conclude the work in Sec. 6.

2 Background

2.1 Prevalence of Metabolic Syndrome

As mentioned above, metabolic syndrome has been generally regarded as one of the most serious diseases over the world. It brings not only the problem of public health, but also the problem of economical cost. It is known that many countries try to prevent/reduce metabolic syndrome, but as the background of the activity there are the problem of high prevalence of metabolic syndrome people in many countries [2].

In United States, the metabolic syndrome prevalence is over 20% and about 47 million U.S. residents have the metabolic syndrome [2]. In Britain, National Diet & Nutrition Survey [3] reported that the number of patients is 9.4 million and about 10.2 million people are in the spare group. In Japan, The National Nutrition Survey reports (2008) [4] that the number of patients are estimated at 25.3% and spare group people at 21.9% in male, and 10.6% and 8.3% in female, respectively. Note that the considerable part of the patients is in working age. Further, the prevalence of diabetes, into which metabolic syndrome may grow when it gets significant, are also reported high in both U.S. and Japan [5][6].

2.2 Activities to Reduce Metabolic Syndrome

To reduce or prevent metabolic syndrome, various activities are going on in several countries. Specifically in Japan, the government has established laws in which companies are responsible to make efforts to reduce weighted staff, and are executed from April 2008. Consequently, many companies started to pay for the effort to reduce the level of metabolic syndrome of their employees, not to be penalized for it. As a result, now it is commonly seen that healthcare nurses continuously make instruction to improve health of the employees of companies.

Now we focus on how the healthcare nurses make instruction to people. In fact, the main method that the health care nurses apply in order to improve health of the metabolic syndrome candidates is the decision of small daily goals that they try to pursuit every day at their home. Note that the daily goals are determined individually under instruction of

a healthcare nurse not too difficult to pursuit. In many cases, the daily goals are selected and many of them are intended to do daily moderate exercise or improve nutrition balance of daily meals. But in practice, many people in fact are not able to continue to pursuit the goals every day since it is difficult to keep their willing to continue for such personal home activities. The problem is that many people tend to stop such activities before long, which significantly degrades the performance of the instruction of healthcare nurses.

2.3 Assisting Health Promotion using Information Systems

As one of the ways to make people continue their health promoting activities, there are various web services which intend to have such people informed about or interested in their health promoting activities. [7][8][9] For instance, Matsumoto et al. [7] proposed an information system to assist to improve dietary habit, by recording user's meal history and providing users nutritionally well-balanced menus based on medical information. This system would help users improve their daily nutritional balance without healthcare nurses. As another instance, an web system K-zoku [8] tries to help users build their health promoting plans and visualize the daily achievements to motivate users to continue diet activities. In K-zoku users can determine their daily goals under the guide of provided medical information, and record daily achievements. Users always check their achievement level of their activities to motivate to continue their plans. Those systems are useful to help users promote their health, but healthcare nurses do not related with the system. Since users have to try to lose weight alone without professional knowledge, the effect of this kind of systems is limited.

On the other side, several information systems exist which try to help healthcare nurses instructing users health promoting [10] [11] [12]. This kind of systems typically try to help communications via the Internet between healthcare nurses and users in home who are trying to achieve daily goals i.e., users input their daily achievement in pursuiting goals into the system. then their healthcare nurses check them and provide advices periodically. There are several success cases of this kind of systems in Japan and consequently it is now regarded that the advices from healthcare nurses are actually effective to motivate users to continue their health promoting activities. For example, Yoshihiro et al. [10] proposed a system to support healthcare activity of (relatively light) diabetes people. In this system users input daily cure records into the system via mobile phones and nurses send back advices periodically. As another example, Harasuma Diet by Hitachi co.ltd. [11] is a commercial web system which supports healthcare nurses instructing users health promoting. In this system "100 kcal card," which shows a small goal to reduce energy of 100 kilo calories, is prepared to help instruction, and once the daily goals to try are determined, users records the daily achievements of the goals. By using 100 kcal cards, the labor of healthcare nurses to determine users' daily goals is considerably reduced and also periodical advice improves the health promotion effect of the users.

However, in fact, the effect to keep users' motivation to continue their health promoting activity is still limited so that considerable part of the people quit their activities. One solution for this problem would be that healthcare nurses pay more effort to communicate with users, but it is not practical. To improve the health promoting effects without increasing healthcare nurses, some systems introduces a mechanisms to promote communication among users using the mechanisms of SNS (Social Network Services) [13]-[17], i.e., they allow users to send messages to other users. This mechanism actually increase user's communication, however, the number of users who send messages are quite limited since sending message is so time consuming.

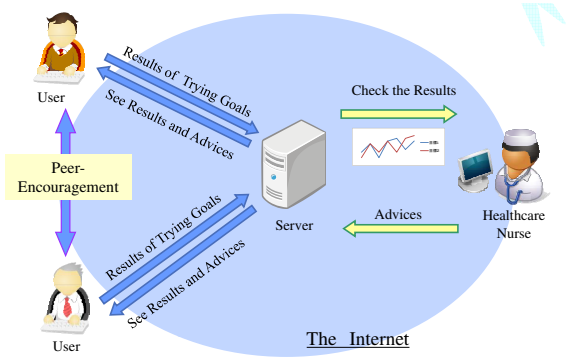


Figure 1: Overview of Our System Scenario

Thus in this paper, to help healthcare nurses to assist users' health promotion activities, we propose a mechanism which enable users to encourage one another among them (which we call peer-encouraging mechanisms) with simple one-click action. By this peer-encouraging mechanism, users will keep their motivation by themselves without increasing labor of healthcare nurses.

3 Proposed Mechanisms

3.1 The Concept

We firstly describe the base system structure that we intend to introduce the peer-encouragement mechanisms. Figure 1 shows the basic functions of such systems. In the Internet there is a web server in which our server programs are installed. Similar to the typical healthcare supporting system, it is expected that healthcare nurses consult users to determine his daily goals to achieve, and the daily goals are set into the server. Suppose that users input their daily achievements into the web server everyday. Then his healthcare nurses are able to check the achievement history from the hospital, and to send some advices if needed.

In this situation, the proposed mechanisms called "peer-encouragement" is done with simple "one-click" operation among users. An user of our system checks other users' daily achievements, and if he feel like encouraging one, then he make an encouragement by one-click action to raise motivation of other users. Note that we prepare two sort of encouragement, i.e., "encouragement" and "admire" that we can select case by case (both of those are called "encouragement" in the following.) Note that we do not think that only this simple mechanism would make effect. Several environmental

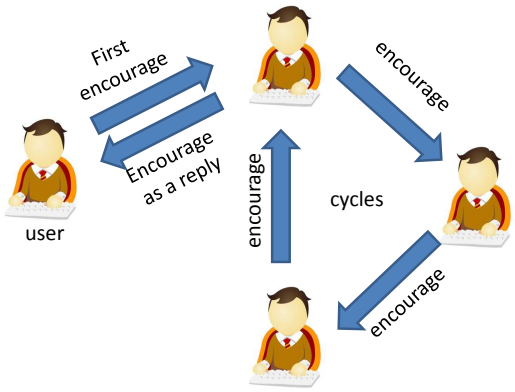


Figure 2: Cycles of Encouragements among Users

mechanisms which work together with this one-click action is essential.

Our idea is to create cycle of encouragements among users. If a user is encouraged by others, the user's motivation would be raised a little, and then he not only goes to continue his health promoting activities, but also he would feel like encouraging other users. This phenomenon makes a chain of encouragements, and it grows to be a cycle after all. In this way, we expect a sustainable cycles among users which always raise their motivation to continue their activities. Figure 2 illustrates the phenomenon: once an user first "encourage" someone, then he not only encourages the user as a reply, but also encourages others in turn. Chain of this phenomenon would create cycles. In this way the first single encouragement would be augmented to be circulated among users.

To create cycles stably, several environmental mechanisms are essential. First, the encouragement operation should be sufficiently simple and clear so that users can make encouragements without labor and stress. Next, sufficient number of encouragements is necessary to keep stable cycles of encouragements. Finally, the encouragements should result in some kind of connectivity between users so that users can compare themselves with other users. Over those points we will discuss in the next subsections.

3.2 Classifying Daily Goals

Now we consider about the scene that an user makes encouragement to other users. To create encouragement chain as many as possible, the interface should be simple and clear not to have users feel labor or stress. Therefore, we propose "one-click encouragement," which is done by simply pressing "encourage" button in the screen. However, not only simple enough the interface is, but also the interface should be clear such that we can imagine the healthcare activity of the user enough to feel like encouraging him. Usually, the healthcare activities of the users are imagined from their achievement history (thus healthcare nurses are able to send proper advices in the typical systems). But note that there are the problem of privacy when users see the other user's daily achievement history, i.e., many people would not feel well when their achievement history is put open in public. Especially, the specific description of daily goals may not accepted by some people.

On this problem, our solution is to classify daily goals into several categories. Then, when we see another user's history, we are able to understand the sort of his each daily goal (without seeing the specific description of daily goals). This would include acceptable information for us to imagine his activities and come to feel like encouraging him. As the categories for the purpose above, we select the following 8 categories:

Food Amount (FA): Daily goals in this category try to reduce daily calories ingested to reduce weight, e.g., stop eating between meals, eat 2/3 of usual amount of rice, etc.

Nutrition (NT): This category tries to improve nutritious balance of daily foods, e.g., eat fruits 3 times a week, drink coffee without sugar, eat vegetable salad once daily, etc.

Light Exercise (LE): This category tries to perform light exercise, e.g., use stairs instead of elevator, perform stretching after taking a bath, etc.

Hard Exercise (HE): This category tries to perform hard exercise, e.g., walk 10,000 steps a day, jog outside everyday, perform physical training 5 times a week, etc.

Daily Measurements (DM): This category tries to measure their body periodically, e.g., measure weights twice (morning and evening) a day, etc.

Prohibit Drinking (PD): This category tries to quit drinking alcohol, e.g., quit drinking twice a week, drink 1/2 of the usual amount of beer, etc.

Prohibit Smoking (PS): This category tries to quit smoking, e.g., quit smoking 3 days a week, reduce the number of cigarettes to 1/2 of usual, etc.

Others (OT): This category includes daily goals which are not included in the above 7 categories, e.g., get up early, chew foods sufficiently many times, etc.

3.3 Augmenting the Number of Encouragements

If the probability to occur chain encouragements is low, the number of encouragement reduces as time passes and cycles would disappear soon. For this problem, we propose two solutions. One is to enforce users to encourage other users when they input daily records. This method provides a constant number of encouragements everyday not to reduce cycles. An anxiety is that users feel the enforced encourage operation laborious or stress, resulting in stopping to use our system. This point should be confirmed in the evaluation.

Another solution is to create automated encouragements to augment the number of encouragements. This also provide encouragements everyday not to reduce cycles. The anxiety of this method is that users would be discouraged when they know that computers, but not real people, are generating encouragements, resulting in lose interests to use this system. In this work we selected the former solution and developed a prototype system based on it.

Table 1: Software Used for the Server

Software	Version
CentOS	4.3
Apache	2.0.52
PostgreSQL	7.4.13
PHP	4.3.9
Postfix	2.2.2.10

3.4 Connection among Users

To create sustainable cycles of encouragements, we regard it also important to create connection among users. The connection among users here means formation of some recognition against other users, i.e., feeling of friends or rivals who lead us to act with or compare us with those people. Through those connections we intend to make users interested in other individual users, by which we expect to augment users' motivation for their activities and the number of encouragements. Specifically, if the list of users who recently encouraged you is given, one would go to see the daily activity of the users, and then he would make encouragements for them with high probability as a reply for them. As another example, if a ranking of the number of encouragements is given, one would check users around the rank of you. Further, if one found users whose daily activities are similar to him/her, then he/she may wish to continuously watch them as his/her rival user.

We regard such kind of connections among users as quite essential when discussing the formation of sustainable cycles. Consequently, in our system design, we implement the following as shown afterwards:

- the list of users who recently encouraged you,
- the list of users you recently encouraged,
- the list of users you want to watch (watch list),
- and ranking of 1-week achievement score and the number of encouragements given.

Also as information by which one may be interested in others, we allow users to input their profile information and put it open to public.

4 System Implementation

4.1 overview

We carefully designed the system with the peer-encouragement mechanism and the environmental mechanisms in order to make the circulation of encouragements work well. Overview of the system has been explained in Sec. 3.1 and Fig. 1. We implemented the server using the software shown in table 1. Note that, since in this paper we evaluate the effect of peer-encouragement mechanisms, we illustrate the interfaces for general users only (i.e., the interface for healthcare nurses are not included in this paper.)

Figure 3: Top View

The interfaces for general users consist of 7 views: top view, daily goals settings, achievement history view, peer-encouragement view, ranking view, profile edit view, and settings. We omit the transition diagram since users can move among those views using the menu bar seen in the upper part of each view.

4.2 Top View

Fig. 3 shows the top view of the system, to which users first come after the authentication. Three fields (1)(2) and (3) are shown, which is to be explained below. The left button of field (1) brings users to the pop-up window in which users input their daily achievements for daily goals. In our system, users can input the achievement of each daily goal, body weight, and a short comment of the day. After the input of those daily records, users are brought to the peer-encouragement view to enforce them to make encouragement for other users. The right button in field (1) brings users to the peer-encouragement view directly. The detail of the peer-encouragement view is seen in Sec.4.5.

The field (2) of Fig. 3 shows the 1-week daily achievement history of an user. In this figure, three goals are shown by categories LE(Light Exercise), NT(Nutrition) and FA(Food Amount). Each goal's 1-week achievement history is seen in a row. Each row consists of two sub-rows. The upper sub-row shows the daily achievement of the daily goal in four levels, i.e., double O, O, triangle and X. If the goal is weekly goal (i.e., the goal to achieve the fixed number in a week, e.g., twice a week), different character meaning "excuse" may be used, which means the goal is not achieved but is allowed. The lower sub-row shows the number of encouragements received from other users in each day. This count is done for each kind of encouragements "encouragement" and "admire" represented by different color of stars. In the second right-

Figure 4: Daily Goals Setting View

most column, the statics values including the number of days continuing to input achievement are shown. In the rightmost column, the rankings of achievement score in the past 1-week in each goal category are shown. The second lowest row shows the daily comments of the user. In the lowest row body weights are seen if they are input by the user.

The field (3) of Fig. 3 shows the three *user lists*; *the watch list* is the list of users you want to watch continuously, *the encourage user list* is the list of users you encouraged recently, and *the encouraged user list* is the list of users who encouraged you recently. Note that for each user the list of categories of his daily goals and the newest comment is shown. By clicking each user, we are brought to the top view of the user. The design of the top view of other users consist of only the field (2) and the function that you make encouragements for the user is added, i.e., the two encouragement buttons are added.

4.3 Daily Goals Setting View

Fig. 4 shows the daily goals setting view, in which at most 10 daily goals are set. There are seven fields to explain. Field (1) shows the examples of goals for reference. To guide users the correspondence between text-form daily goals and the categories, this field is necessary. The other six fields are the part of the web form to set daily goals. In the second leftmost column (field (2)) we fill the text representation of daily goals. Then, in the leftmost column (field (3)) we select the corresponding category among 8 candidates shown in Sec. 3.2. In the column (4), we select the type of the goal, i.e., daily goal or weekly goal (if we select weekly goal, we can select the number of expected achievements in a week.) In the column (5) users select whether the goal is open in public or not; users can select at most 3 goals to be open, to which other users make encouragements. In the column (6) we see the period of time that we are trying to achieve that daily goal, and users can reset the starting day and restart for this goal by pressing the button (i.e., the achievement history is cleared). In the column (7) we see the ratio of days that the user inputs the achievements so far.

4.4 Achievement History View

In the achievement history view, users can see the daily achievement history of themselves. The items that users can see is all the input data they input so far, the record of encouragements received from other users so far, and some related statistic values. We omit the explanation of this view since there is no specific feature in this view.

4.5 Peer-encouragement View

In the peer-encouragement view we can encourage other users with one-click “encouragement” or “admire” operation. To augment the number of encouragements, we make users encourage other users at least once a day, as mentioned in Sec.3.3. So, after inputting the daily achievement, users are forced to come to this view and to make encouragements. Also, users can visit this view at their will after that.

Fig. 5 shows the view to encourage other users, where 5 users are selected and shown in 5 rows. Field (1) of Fig. 5 shows the 5 users whose nickname is hidden. Those users are selected basically at random, but we embed a mechanism that the users of less received encouragement is more likely to be selected. Field (2) shows the achievement history of users as a reference data to decide whether you encourage them or not. In this field, the information seen is mostly the same as the field (2) of top view (Fig. 3), i.e., the 1-week achievement history of three daily goals, the daily comments, and the 1-week history of received encouragements for each user. In the field (3) we see some statistic variables which is the same as the field (2) of Fig. 3, i.e., the length of time in days since the users start trying their achievement, the number of input in days, and the number of each level of achievements (i.e., double O, O, triangle, and X). You can encourage those 5 users according to the information described above; the encouragement is done by two buttons shown in field (4) in Fig. 5. Each of two buttons corresponds to two sorts of encouragements, i.e., “encouragement” and “admire.” Note that you are not necessary to encourage all those 5 people. However, at least one encouragement is required to finish this view. When you finish encouraging, you can continue to encourage other users or return to your top view, using the buttons seen in field (5).

4.6 Ranking View

In the ranking view we can see the weekly ranking of achievement scores and the number of encouragements obtained (for each of “encourage” and “admire”). The ranking is created for each category of daily goals; each ranking consists of users who have at least one daily goal which belong to that category. Fig. 6 shows the ranking view. In field (1) we can select the ranking to look: the upper tabs correspond to 8 categories and “all categories” added at the leftmost tab. Specifically, from its left side, the categories All, FA, NT, LE, HE, DM, PD, PS, OT described in Sec. 3.2 are shown. If we click one, the ranking shown in fields (3)(4)(5) changes to the corresponding ranking table. The three buttons in field (1) allows

Figure 5: Peer-encouragement View

Figure 6: Ranking View

us to select the sort of ranking, i.e., from the left side, achievement score, the number of "admire" obtained, and that of "encouragement" are placed. The field (2) shows your rank in the ranking table selected in the operation of field (1). In fields (3)(4)(5) the ranking table is shown. The leftmost column shows the user nicknames and their rank, The second leftmost column shows the users' three goal categories and the 1-week achievement history. The next three columns show the achievement score, the number of "admire," and that of "encouragement," respectively. The achievement score is simply calculated as the weighted sum of daily achievement levels of a user. Specifically in this system, double O is 4 points, O is 3 points, triangle is 2 points, and X is 1 point. The newest daily comments of each user are also seen in those fields. Note that fields (3)(4) and (5) are dyed with different colors: in the ranking you can easily recognize yourself by orange color (4), do the users in your watch list by blue color (3), and other users are not colored (white) like field (5). When we click the users seen, then we move to their top view where we can make encouragements for them.

4.7 Profile Edit View

In this system users can input their personal profiles into the system and open them in public. The data items which are possible to input are: living area, sex, birthday, blood type, hobby, profession, and introduction texts. The profile is intended to be the information to connect users, i.e., an user is possibly interested in other users from the information. This information can be seen in their top view, and also we can search users using those items from the search form of the system.

4.8 Setting View

In the setting, users can modify the settings of remainder mails, which is sent everyday at the determined time not to forget inputting daily achievements. Users can select whether they use this function, and can set the e-mail address and the time to receive the e-mail. In this e-mail, information of receiving encouragements are seen so that they can get to know someone has encouraged them.

5 Evaluation

5.1 Methods

We evaluated the one-click peer-encouragement mechanism using the system implementation introduced in the previous section. We operate the system about one month with 29 people and obtain event logs and answers of questionnaire. As for the questionnaire, the selected questions and the answers are shown in Table 2. Here, we adopt 5-level rating for all the questions, where 5 is the most positive and 1 is the most negative evaluation. In this experiment, we mainly try to evaluate how effectively the peer-encouragements are circulating in advance of practical use in the scene of healthcare instructions.

Several conditions, however, should be properly prepared to obtain reliable results from the experiments. Especially, whether the daily goals are adequately determined or not affects significantly on the evaluation results, so we conduct that each participant consults a healthcare nurse to determine their adequate daily goals. Although many of participants are not even candidates of metabolic syndrome, they all have room to improve their lifestyle, so that they all set daily goals with their will to improve their lifestyle and promote their health.

5.2 Results

First we present the results on circulation of encouragements observed. Fig. 7 shows the number of encouragements observed, presented as accumulated numbers in time course. Here, the "enforced" encouragements mean the encouragements which are made when users first come to the peer-encouragement view without their will as described in Sec.4.5., while the "active" encouragements mean those which are made

Table 2: Questions and the Results (Selected from the questionnaire)

No.	Questions	Evaluation (# of answers)					
		average	5	4	3	2	1
A-1	How much did you motivated to continue your activities from the received encouragements?	3.5	1	14	3	0	3
B-5	Did you feel laborious to make encouragements?	3.5	5	6	3	5	1
F-1	Did you keep your motivation to continue your activities throughout the experiment?	4.0	3	16	1	0	1
F-2	Do you think that this system will work in the practical healthcare scenes?	3.8	3	13	3	2	0

when users come to the peer-encouragement view or other users' top view at their will.

From this graph, the number of the active encouragements are far larger than enforced encouragements (about 86% of all the encouragements were the active one). Note that we see the rapid change of the curve around 2010/1/5, which would be the effect of new-year days. Fig. 8 shows the specification of the active encouragements classified by sources from which users made encouragements from. Here, "user lists" means the three user lists (the watch list, list of the users who encouraged you, and the list of users you encouraged) described in Sec.4.2, i.e., they are the encouragements that users click other users in those user lists and encourage them in their top view. The "encouragement view" means the encouragements from peer-encouragement view described in Sec.4.5. The "ranking view" means the encouragements from ranking view described in Sec.4.6, i.e., users click other users in the ranking view and made encouragements at their top view. The "user's top" means the encouragements made at users' top view to which users come using keyword search form of the system.

The graph shows that the encouragements from the user lists and encouragement view count large number. Consequently, users are not only interested in encouragement view where users are randomly seen, but also encourage others using some kind of "connections" among users described in Sec.3.4. Fig. 9 shows further specification of "user lists" of Fig. 8. From the graph, the watch list does not work as a source of encouragements after 10 days past, while encouragements from other two user lists continuously increase in the whole period. This result implies that users tend to make encouragements based on the past encouragements of themselves or other users, e.g., they tend to make encouragements as a reply for other users, or tend to re-check the users they have encouraged before.

Note that, from the result of the question B-5 seen in Table 2, users feel everyday's enforced encouragements somewhat laborious and time consuming. However, on the other side, they still continue to make encouragements at their will not only to the users from whom they received encouragements, but also to the other users from peer-encouragement view. This implies that the labor for enforced encouragements would be at allowable level and the motivation to continue which comes from receiving encouragements would be larger

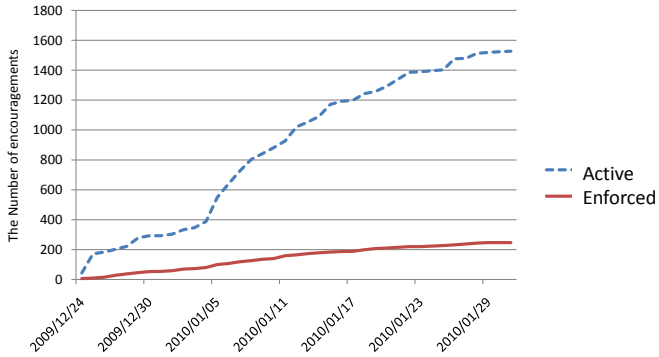


Figure 7: Number of Active/Enforced Encouragements

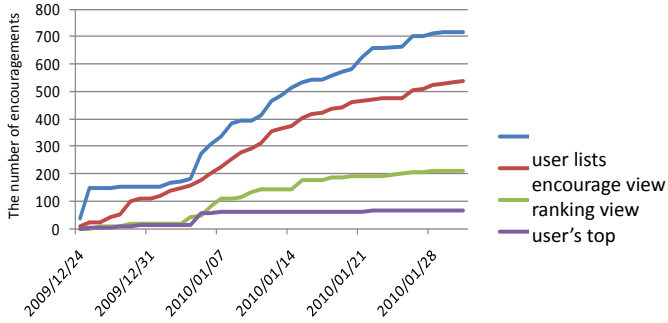


Figure 8: Source of Active Encouragements

in total.

As seen above, in this experiment new encouragements are constantly generated by enforced encouragements mechanisms, and the encouragements actually chain via peer-encouragement view or via the mechanisms to keep "connections" such as user lists. As a result, encouragements are continuously supplied by users in the whole period of the experiment. Consequently, it is concluded that our mechanisms actually circulate encouragements among users.

Next we will see the effects of encouragements on users' motivation to continue their activity. From the result of question A-1 (see Table 2), high ratio of users answer that receiving encouragements from other users raise their motivation to continue their activities. Note that, as for 3 users who give the rating 1 for question A-1, we found that they also give low rating (1 or 2) for B-5, i.e., it infers that they are not

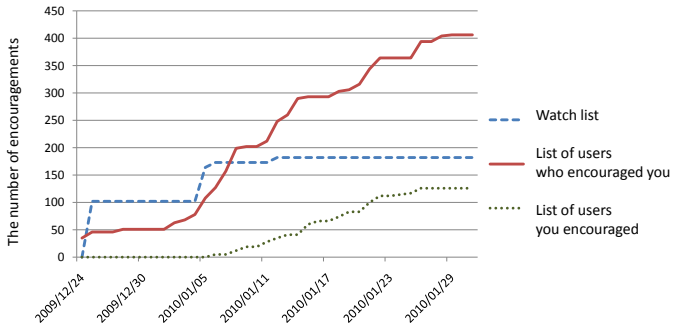


Figure 9: Source of Active Encouragements made from User Lists

motivated since they feel too much labor or stress in the operation of making encouragements. On the other side, from the event log, the number of daily records and that of received encouragements have very high correlation coefficient 0.85. Also the number of created encouragements and that of received ones have also high correlation coefficient 0.65. Consequently, users are typically motivated by receiving encouragements (except for few exceptions seen in the result of A-1), and in fact the number of users' daily records has deep relation with encouragements. Those results clearly show that the encouragements have motivated users to continue their activities.

Finally we see the total impression of users for our system. Question F-1 (in Table 2) shows that most of the users could keep their motivation throughout the period of the experiment. Also, F-2 shows that many users answer that they have good impressions and possible to have active impression in putting our system into practice. Although they are not professional in this area, but this result shows that at least many people feel well from the user's point of view.

6 Conclusions

In this paper we proposed a peer-encouragement mechanisms to improve users motivation to continue their healthcare activity while the amount of healthcare nurses' labor does not increase. We present our concept of peer-encouragement mechanisms in the healthcare scenes, and show the system design of a healthcare supporting systems to work in practice. Through one-month experiment, we confirmed that the proposed peer-encouragement mechanisms work effectively to create circulation of encouragements among users, and consequently the users are continually motivated to continue their healthcare activity without increasing healthcare nurse's work.

For the future, evaluation in the practical scene is necessary to try the mechanism to be put in practice. We would like to improve our system in consideration of the results obtained from this experiment, and try to conduct experiments in practical scenes.

REFERENCES

[1] S.M. Grundy, Metabolic Syndrome Pandemic, Arteriosclerosis, Thrombosis, and Vascular Biology, Vol.28, pp.629-636, 2008.

[2] E.S. Ford, W.H. Giles, W.H. Dietz, Prevalence of the Metabolic Syndrome Among US Adults: Findings From the Third National Health and Nutrition Examination Survey, JAMA 287, pp.356-359, 2002.

[3] Office for National Statistics, et al., The National Diet & Nutrition Survey: adults aged 19 to 64 years, 2004.

[4] Ministry of Health, Labour and Welfare of Japan, The National Nutrition Survey, Japan: NNS-J, 2008 (In Japanese).

[5] National Diabetes Fact Sheet, Department of Health and Human Services of U.S.A, 2005.

[6] Fact-finding Survey on Diabetes in Japan, Ministry of Health, Labor and Welfare of Japan, 2002 (in Japanese).

[7] T. Matsumoto, Y. Shimada, T. Teo, S. Kawaji, A Design of Information System Improving Dietary Habit Based on Individual Clinical Data and Life Style, 21st IEEE International Symposium on Computer-Based Medical Systems(CBMS2008), pp.173-175, 2008.

[8] K-zoku, http://diet.kzoku.jp/ (in Japanese).

[9] Tekuteku walking link, http://www.health-walking.com/ (in Japanese).

[10] T. Yoshihiro, K. Ikemoto, K. Mori, S. Kagawa, Y. Yamamoto, and M. Nakagawa, System support of Diabetic Self-treatments Using Mobile Phones, Proceeding of the 2006 Conference on Knowledge-Based Software Engineering (JCKBSE2006), pp. 263-272, 2006.

[11] Harasuma Diet, http://www.hitachi.co.jp/Prod/comp/app/harasuma/index.html (in Japanese).

[12] Medicom, http://www.medicom.sanyo.com/smile/ (in Japanese).

[13] Kenko-Seikatsu Navi, http://kenkonavi.val.co.jp/ (in Japanese).

[14] Diet Social Network, http://dietsocialnetwork.com/ .

[15] Dieting and weight loss, http://www.diet.com/ .

[16] Extrapounds.com: weight loss social network, http://www.extrapounds.com/ .

[17] Live Health Club, http://www.livehealthclub.com/.

Reachability Analysis of Probabilistic Real-Time Systems Based on CEGAR for Timed Automata

Takeshi NAGAOKA†, Akihiko ITO†, Toshiaki TANAKA†, Kozo OKANO†, and Shinji KUSUMOTO†

†Graduate School of Information Science and Technology, Osaka University
Yamadaoka 1–5, Suita City, Osaka, 565-0871, Japan

Abstract - Model checking techniques are considered as promising techniques for verification of information systems due to their ability of exhaustive checking. Well-known state explosion, however, might occur in model checking of large systems. Such explosion severely limits the scalability of model checking. In order to avoid it, several abstraction techniques have been proposed. Some of them are based on CounterExample-Guided Abstraction Refinement (CEGAR) technique proposed by E. Clarke *et al.*.

This paper proposes a reachability analysis technique for probabilistic timed automata. In the technique, we abstract time attributes of probabilistic timed automata by applying our abstraction refinement technique for timed automata proposed in our previous work. Then, we apply probabilistic model checking to the generated abstract model which is just a markov decision process (MDP) with no time attributes. This paper also provides some experimental results on applying our method to IEEE 1394, FireWire protocol. Experimental results show our algorithm can reduce the number of states and total execution time dramatically compared to one of existing approaches.

Keywords: Probabilistic Timed Automaton, CEGAR, Model Checking, Real-time System, Formal Verification

1 Introduction

Model checking[1] techniques are considered as promising techniques for verification of information systems due to their ability of exhaustive checking. For verification of real-time systems such as embedded systems, timed automata are often used. On the other hand, probabilistic model checking[2]–[4] can evaluate performance, dependability and stability of information processing systems with random behaviors. In recent years, probabilistic models with real-time behaviors, called probabilistic timed automata (PTA) attract attentions. As well as traditional model checking techniques, however, state explosion is thought to be a major hurdle for verification of probabilistic timed automata.

Clarke *et al.* proposed an abstraction technique called CEGAR (CounterExample-Guided Abstraction Refinement)[5] shown in Fig.1. In the CEGAR technique, we use a counter example (CE) produced by a model checker as a guide to refine abstracted models. A general CEGAR technique consists of several steps. First, it abstracts the original model (the obtained model is called abstract model) and performs model checking on the abstract model. Next, if a CE is found, it checks whether the CE is feasible on the concrete model or not. If the CE is spurious, it refines the abstract model. The last step is repeated until the valid output is obtained. In the CEGAR loop, an abstract model must satisfy the following property: if the abstract model satisfies a given specification, the concrete model also satisfies it.

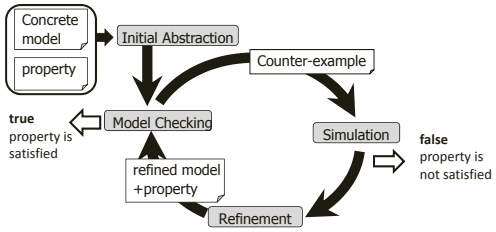


Figure 1: A General CEGAR Technique

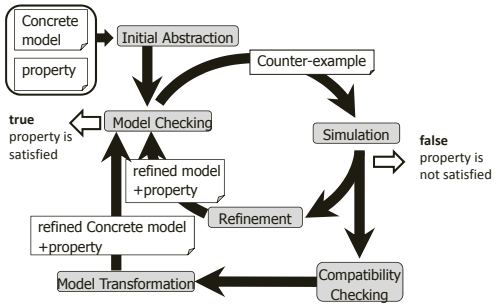


Figure 2: Our CEGAR Technique for Reachability Analysis of a Probabilistic Timed Automaton

In Paper[6], we have proposed an abstraction algorithm for timed automata based on CEGAR. In this algorithm, we generate finite transition systems as abstract models where all time attributes are removed. The refinement modifies the transition relations of the abstract model so that the model behaves correctly even if we don't consider the clock constraints.

This paper proposes a reachability analysis technique for probabilistic timed automata. In the technique, we abstract time attributes of probabilistic timed automata by applying our abstraction technique for timed automata proposed in Paper[6]. Then, we apply probabilistic model checking to the generated abstract model which is just a markov decision process (MDP) with no time attributes. The probabilistic model checking algorithm calculates summation of occurrence probability of all paths which reach to a target state for reachability analysis. For probabilistic timed automata, however, we have to consider required clock constraints for such paths, and choose the paths whose required constraints are compatible. Since our abstract model does not consider the clock

constraints, we add a new flow where we check whether all paths used for probability calculation are compatible. Also, if they are not compatible, we transform the model so that we do not accept such incompatible paths simultaneously. The proposed procedure for the probabilistic timed automata is shown in Fig.2.

This paper also provides some experimental results on applying our method to some examples. Experimental results show our algorithm can reduce the number of states and total execution time dramatically compared to one of existing approaches.

Several papers including Paper[3] have proposed probabilistic model checking algorithms. These algorithms, however, don't provide CEs when properties are not satisfied. Our proposed method provides a CE as a set of paths based on k-shortest paths search. This is a major contribution of our method. The proposed method also performs model checking considering compatibility problem. Few approaches resolve the compatibility problem. Our approach also shows the efficiency via performing experiments.

The organization of the rest paper is as follows. Sec.2 provides some definitions and lemmas as preliminaries. Sec.3 describes our proposed abstraction technique for the probabilistic timed automaton. Sec.4 gives some experimental results. Finally, Sec.5 concludes the paper and gives future works.

2 Preliminary

This section gives some definitions about models used in this paper and also describes a general CEGAR technique.

2.1 Clock and Zone

Let C be a finite set of clock variables which take non-negative real values ($\mathbb{R}_{\geq 0}$). A map $\nu : C \to \mathbb{R}_{\geq 0}$ is called a clock assignment. The set of all clock assignments is denoted by $\mathbb{R}^C_{\geq 0}$. For any $\nu \in \mathbb{R}^C_{\geq 0}$ and $d \in \mathbb{R}_{\geq 0}$ we use $(\nu + d)$ to denote the clock assignment defined as $(\nu+d)(x) = \nu(x)+d$ for all $x \in C$. Also, we use $r(\nu)$ to denote the clock assignment obtained from ν by resetting all of the clocks in $r \subseteq C$ to zero.

Definition 2.1 (Differential Inequalities on C)**.** Syntax and semantics of a differential inequality E on a finite set C of clocks is given as follows:
$E ::= x - y \sim a \mid x \sim a$,
where $x, y \in C$, a is a literal of a real number constant, and $\sim \in \{\leq, \geq, <, >\}$. Semantics of a differential inequality is the same as the ordinal inequality.

Definition 2.2 (Clock Constraints on C)**.** Clock constraints $c(C)$ on a finite set C of clocks is defined as follows:
A differential inequality in on C is an element of $c(C)$.
Let in_1 and in_2 be elements of $c(C)$, $in_1 \wedge in_2$ is also an element of $c(C)$.

A zone $D \in c(C)$ is described as a product of finite differential inequalities on clock set C, which represents a set of clock assignments that satisfy all the inequalities. In this paper, we treat a zone D as a set of clock assignments $\nu \in \mathbb{R}^C_{\geq 0}$

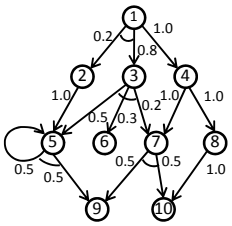


Figure 3: An Example of an MDP

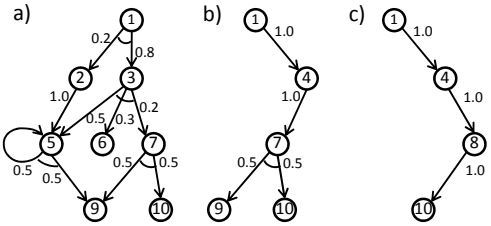


Figure 4: Examples of Adversaries

(For a zone D, $\nu \in D$ means the assignment ν satisfies all the inequalities in D).

2.2 Probability Distribution

A discrete probability distribution on a finite set Q is given as the function $\mu : Q \to [0, 1]$ such that $\Sigma_{q \in Q}\mu(q) = 1$. Also, $support(\mu)$ is a subset of Q such that $\forall q \in support(\mu).\mu(q) > 0$ holds.

2.3 Markov Decision Process

A Markov Decision Process (MDP)[7] is a markov chain with non-deterministic choices.

Definition 2.3 (Markov Decision Process)**.** A markov decision process MDP is 3-tuple $(S, s_0, Steps)$, where
S : a finite set of states;
$s_0 \in S$: an initial state; and
$Steps \subseteq S \times A \times Dist(S)$: a probabilistic transition relation where $Dist(S)$ is a probability distribution over S.

In our reachability analysis procedure, we transform a given PTA into a finite MDP, and perform probabilistic verification based on the Value Iteration[8] technique.

Figure 3 shows an example of an MDP. In the figure, probability distributions are associated with transitions. In the figure, transitions which belong to the same distribution are connected with a small arc at their source points. The MDP has several non-deterministic choices at the state 1 and 4. For example, at the state 1, we have two choices; 1) the control moves to the state 2 with the probability 0.2 and to the state 3 with the probability 0.8, 2) the control moves to the state 4 with the probability 1.0.

2.3.1 Adversary

An MDP has non-deterministic transitions called action. To resolve the non-determinism, an adversary is used. The adver-

sary requires a finite path on an MDP, and decides a transition to be chosen at the next step.

Figure 4 shows examples of resolving the non-determinism of the MDP shown in Fig.3 by some adversaries. Figure 4. a) is the case where we choose the action which moves to the state 2 or state 3 at the initial state 1. On the other hand, b) and c) are the cases where we choose the action which moves to the state 4 at the initial state 1. In the case of b), we choose the action which moves to the state 7 when we move from the state 1 to state 4. Also, in the case of c), we choose the action which moves to the state 8 in the same trace.

Here, if we want to obtain the reachability probability from the state 1 to the state 10, under the adversary of a), we can obtain the probability 0.08 ($= 0.8 \times 0.2 \times 0.5$), which is the minimum reachability probability. On the other hand, under the adversary of c), we can obtain the probability 1.0 ($= 1.0 \times 1.0 \times 1.0$), which is the maximum reachability probability.

2.3.2 Value Iteration

A representative technique of model checking for an MDP is Value Iteration[8]. The Value Iteration technique can obtain both of maximum and minimum probabilities of reachability and safety properties, respectively. At each state, Value Iteration can select an appropriate action according to the property to be checked. Therefore, the technique can produce the adversary as well as the probability.

2.4 Timed Automaton

Definition 2.4 (Timed Automaton)**.** A timed automaton $\mathscr{A}$ is a 6-tuple (A, L, l_0, C, I, T), where
A: a finite set of actions;
L: a finite set of locations;
$l_0 \in L$: an initial location;
C: a finite set of clocks;
$I \subset (L \to c(C))$: a mapping from locations to clock constraints, called a location invariant; and
$T \subset L \times A \times c(C) \times \mathscr{R} \times L$,
where $c(C)$ is a clock constraint, called guards;
$\mathscr{R} = 2^C$: a set of clocks to reset.

A transition $t = (l_1, a, g, r, l_2) \in T$ is denoted by $l_1 \xrightarrow{a,g,r} l_2$. A map $\nu : C \to \mathbb{R}_{\geq 0}$ is called a clock assignment. We define $(\nu + d)(x) = \nu(x) + d$ for $d \in \mathbb{R}_{\geq 0}$. $r(\nu) = \nu[x \mapsto 0], x \in r$, where $\nu[x \mapsto 0]$ means the valuation that maps x into zero, is also defined for $r \in 2^C$.

Definition 2.5 (Semantics of a Timed Automaton)**.** Given a timed automaton $\mathscr{A} = (A, L, l_0, C, I, T)$, let $S \subseteq L \times \mathbb{R}^C_{\geq 0}$ be a set of whole states of $\mathscr{A}$. The initial state of $\mathscr{A}$ shall be given as $(l_0, 0^C) \in S$.
For a transition $l_1 \xrightarrow{a,g,r} l_2 \ \ (\in T)$, the following two transitions are semantically defined. The former one is called an action transition, while the latter one is called a delay transition.

$$\frac{l_1 \xrightarrow{a,g,r} l_2, g(\nu), I(l_2)(r(\nu))}{(l_1, \nu) \overset{a}{\Rightarrow} (l_2, r(\nu))}, \qquad \frac{\forall d' \leq d \quad I(l_1)(\nu + d')}{(l_1, \nu) \overset{d}{\Rightarrow} (l_1, \nu + d)}$$

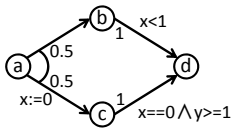


Figure 5: An Example of a PTA

Definition 2.6 (A Semantic Model of a Timed Automaton)**.** For timed automaton $\mathscr{A} = (A, L, l_0, C, I, T)$, an infinite transition system is defined according to the semantics of $\mathscr{A}$, where the model begins with the initial state.

2.5 Probabilistic Timed Automaton

A PTA is a kind of a timed automaton extended with probabilistic behavior. In the PTA, a set of probabilistic distributions is used instead of a set T of discrete transitions on the timed automaton.

Definition 2.7 (Probabilistic Timed Automaton)**.** A probabilistic timed automaton PTA is a 6-tuple $(A, L, l_0, C, I, prov)$, where
A: a finite set of actions;
L: a finite set of locations;
$l_0 \in L$: an initial location;
C: a finite set of clocks;
$I \subset (L \to c(C))$: a location invariant; and
$prob \subseteq L \times A \times c(C) \times Dist(2^C \times L)$: a finite set of probabilistic transition relations, where $c(C)$ represents a guard condition, and $Dist(2^C \times L)$ represents a finite set of probability distributions p. The Distribution $p(r, l) \in Dist(2^C \times L)$ represents the probability of resetting clock variables in r and also moving to the location l;

Figure 5 shows an example of a PTA. In the figure, from the location a, it moves to the location b with the probability 0.5 and also moves to the location c letting the value of the clock x reset to zero with the probability 0.5. Both of the arcs starting location a are connected with a small arc at their source points, which represents that they belong to the same probability distribution.

Definition 2.8 (Transitions of a Probabilistic Timed Automaton)**.** For $PTA = (A, L, l_0, C, I, prov)$, 6-tuple (l, a, g, p, r, l') represents a transition generated by a probabilistic distribution $(l, a, g, p) \in prob$ such that $p(r, l') > 0$.

Definition 2.9 (Semantics of a Probabilistic Timed Automaton)**.** Semantics of a probabilistic timed automaton $PTA = (A, L, l_0, C, I, prob)$ is given as a timed probabilistic system $TPS_{PTA} = (S, s_0, TSteps)$ where,

- $S \subseteq L \times \mathbb{R}^C$;
- $s_0 = (l_0, 0^C)$; and
- $TSteps \subseteq S \times A \cup \mathbb{R}_{\geq 0} \times Dist(S)$ is composed of action transitions and delay transitions.

  a) action transition
  if $a \in A$ and there exists $(l, a, g, p) \in prob$ such

that $g(\nu)$ and $I(l')(r(\nu))$ for all $(r, l') \in support(p)$, $((l, \nu), a, \mu) \in TSteps$ where for all $(l', \nu') \in S$

$$\mu(l', \nu') = \sum_{r \subseteq C \wedge \nu' = r(\nu)} p(r, l').$$

b) delay transition
if $d \in \mathbb{R}_{\geq 0}$, and for all $d' \leq d$, $I(l)(\nu + d')$, $((l, \nu), d, \mu) \in TSteps$ where $\mu(l, \nu + d) = 1$.

In this paper, using a location l and a zone D, we describe a set of semantic states as $(l, D) = \{(l, \nu) \mid \nu \in D\}$.

A probabilistic timed automaton is said to be well-formed if a probabilistic edge can be taken whenever it is enabled[2]. Formally, a probabilistic timed automaton $PTA = (A, L, l_0, C, I, prob)$ is well-formed if

$$\forall (l, g, p) \in prob.\ \forall \nu \in \mathbb{R}^C_{\geq 0}.\ (g(\nu)) \rightarrow \forall (r, l) \in support(p).\ I(l)(r(\nu)).$$

In this paper, we assume that a given PTA is well-formed.

Definition 2.10 (Path on a Timed Probabilistic System)**.** A path ω with length of n on a timed probabilistic system $TPS_{PTA} = (S, s_0, TSteps)$ is denoted as follows.

$$\omega = (l_0, \nu_0) \xrightarrow{d_0, \mu_0} (l_1, \nu_1) \xrightarrow{d_1, \mu_1} \dots \xrightarrow{d_{n-1}, \mu_{n-1}} (l_n, \nu_n)$$

, where $(l_0, \nu_0) = s_0$, $(l_i, \nu_i) \in S$ for $0 \leq i \leq n$ and $((l_i, \nu_i), d_i, \mu) \in TSteps \wedge ((l_i, \nu_i + d_i), 0, \mu_i) \in TSteps \wedge (l_{i+1}, \nu_{i+1}) \in support(\mu_i)$ for $0 \leq i \leq n - 1$.

For model checking of a probabilistic timed automaton, we extract a number of paths and calculate a summation of their occurrence probabilities in order to check the probability of satisfying a given property. The important point is that we have to choose a set of paths which are compatible with respect to time elapsing.

Lemma 2.1 (Compatibility of two paths)**.** If two paths $\omega^\alpha = (l^\alpha_0, \nu^\alpha_0) \xrightarrow{d^\alpha_0, \mu^\alpha_0} (l^\alpha_1, \nu^\alpha_1) \xrightarrow{d^\alpha_1, \mu^\alpha_1} \dots \xrightarrow{d^\alpha_{n-1}, \mu^\alpha_{n-1}} (l^\alpha_n, \nu^\alpha_n)$ and $\omega^\beta = (l^\beta_0, \nu^\beta_0) \xrightarrow{d^\beta_0, \mu^\beta_0} (l^\beta_1, \nu^\beta_1) \xrightarrow{d^\beta_1, \mu^\beta_1} \dots \xrightarrow{d^\beta_{m-1}, \mu^\beta_{m-1}} (l^\beta_m, \nu^\beta_m)$ on a timed probabilistic system TPS_{PTA} satisfy the following predicate $isCompatible$, then ω^α and ω^β are said to be compatible.

$$isCompatible(\omega^\alpha, \omega^\beta) = \begin{cases} \text{true}, & \text{if } \forall i < min(n, m).\ l^\alpha_i = l^\beta_i \wedge d^\alpha_i = d^\beta_i \\ & \text{or there exists } i < min(n, m) \text{ such that} \\ & l^\alpha_i \neq l^\beta_i \wedge d^\alpha_i = d^\beta_i \wedge \\ & \forall j < i.\ l^\alpha_j = l^\beta_j \wedge d^\alpha_j = d^\beta_j \\ \text{false}, & \text{otherwise.} \end{cases}$$

Above predicate $isCompatible$ stands for that two paths are compatible if and only if one path is a prefix of the other, or same amount of delay is executed in both paths at the branching point of them.

Lemma 2.2 (Compatibility of a set of paths)**.** If a set Ω of paths on a timed probabilistic system TPS_{PTA} satisfies the following predicate $isCompatible$, then all of the paths over Ω are said to be compatible.

$$isCompatible(\Omega) = \begin{cases} \text{true}, & \text{if } \forall i \leq min(\Omega) \bigwedge_{\substack{\omega^\alpha, \omega^\beta \in \Omega \\ \wedge \omega^\alpha \neq \omega^\beta}} (l^\alpha_i = l^\beta_i \wedge d^\alpha_i = d^\beta_i) \\ & \text{or there exists } i \leq min(\Omega) \text{ such that} \\ & \bigwedge_{\substack{\omega^\alpha \omega^\beta \in \Omega \\ \wedge \omega^\alpha \neq \omega^\beta}} (l^\alpha_i \neq l^\beta_i \wedge d^\alpha_i = d^\beta_i \wedge \bigwedge_{j \leq i} (l^\alpha_j = l^\beta_j \wedge d^\alpha_j = d^\beta_j)), \\ & \text{and also } \bigwedge_{\substack{\Omega' \in 2^\Omega \wedge \\ \Omega' \neq \Omega \wedge |\Omega'| \leq 2}} isCompatible(\Omega') \\ \text{false}, & \text{otherwise.} \end{cases}$$

In Lemma2.2, we give the predicate $isCompatible$ for a set Ω of paths on a timed probabilistic system. In the lemma, we let paths in Ω be compatible if there is no contradiction with respect to time elapsing at the branching point of all the paths in Ω, and also if the compatibility is kept for every subset of Ω which contains more than two paths.

Next, we give a simple example of a pair of paths which does not satisfy the compatibility. In the Fig.5, paths from the location a to d are given as $\omega^\alpha = (a, x = 0 \wedge y = 0) \xrightarrow{0, 0.5} (b, x = 0 \wedge y = 0) \xrightarrow{0, 1.0} (d, x = 0 \wedge y = 0)$ which reaches to d through b, and $\omega^\beta = (a, x = 0 \wedge y = 0) \xrightarrow{1, 0.5} (c, x = 0 \wedge y = 1) \xrightarrow{0, 1.0} (d, x = 0 \wedge y = 1)$ which reaches to d through c. In the path ω^α, we are required to let delay at the location a be less than one unit of time because of the guarded condition $x < 1$ of the transition between b and d. On the other hand, in the path ω^β, we are required to let delay at a be grater than or equal one unit of time because of the condition $x == 0 \wedge y \geq 1$ of the transition between c and d. Like the path ω^α and ω^β, if the required conditions of time elapsing at the branching point are contradict, we cannot use such paths simultaneously in the probability calculation.

2.6 CounterExample-Guided Abstraction Refinement

2.6.1 General CEGAR Technique

Since model abstraction sometimes over-approximates an original model, we may obtain spurious CEs which are infeasible on the original model. Paper [5] gives an abstraction refinement framework called CEGAR (CounterExample-Guided Abstraction Refinement) (Fig.1).

In the algorithm, at the first step (called Initial Abstraction), it generates an initial abstract model. Next, it performs model checking on the abstract model. In this step, if the model checker reports that the model satisfies a given specification, we can conclude that the original model also satisfies the specification, because the abstract model is an over-approximation of the original model. If the model checker reports that the model does not satisfy the specification, however, we have to check whether the CE detected is spurious or not in the next step (called Simulation). In the Simulation step, if we find that the CE is valid, we stop the loop. Otherwise, we have to refine the abstract model to eliminate the

spurious CE, and repeat these steps until valid output is obtained.

2.6.2 CEGAR Technique for a Timed Automaton

In Paper[6], we have proposed the abstraction refinement technique for a timed automaton based on the framework of CEGAR. In this approach, we remove all the clock attributes from a timed automaton. If a spurious CE is detected by model checking on an abstract model, we transform the transition relation on the abstract model so that the model behaves correctly even if we don't consider the clock constraints. Such transformation obviously represents the difference of behavior caused by the clock attributes. Therefore, the finite number of application of the refinement algorithm enables us to check the given property without the clock attributes. Since our approach does not restore the clock attributes at the refinement step, the abstract model is always a finite transition system without the clock attributes.

3 Proposed Approach

In this section, we will present our abstraction refinement technique for a probabilistic timed automaton. In the technique, we use the abstraction refinement technique for a timed automaton proposed in Paper[6]. In addition, we resolve the compatibility problem shown in Sec.2.5 by performing a backward simulation technique and generating additional location to distinguish the required condition for every incompatible path. Figure 2 shows our abstraction refinement framework. As shown in the figure, we add another flow where we resolve the compatibility problem.

Our abstraction requires a probabilistic timed automaton PTA and a property to be checked as its inputs. The property is limited by the PCTL formula $P_{<p}[\text{true } \mathbf{U}\ err]$. The formula represents a property that the probability of reaching to states where err (which means an error condition in general) is satisfied, is less than p.

3.1 Initial Abstraction

The initial abstraction removes all the clock attributes from a given probabilistic timed automaton as well as the technique in Paper[6]. The generated abstract model over-approximates the original probabilistic timed automaton. Also, the abstract model is just an MDP without time attributes.

Definition 3.1 (Abstract Model)**.** For a given probabilistic timed automaton $PTA = (A, L, l_0, C, I, prob)$, a markov decision process $\hat{MDP}_{PTA} = (\hat{S}, \hat{s}_0, \hat{Steps})$ is produced as its abstract model, where

- $\hat{S} = L$
- $\hat{s}_0 = l_0$
- $\hat{Steps} = \{\ (s, a, p) \mid (s, a, g, p) \in prob\ \}$

Figure 6 shows an initial abstract model for the PTA shown in Fig.5 As shown in the figure, the abstract model is just an MDP where all of the clock constraints are removed though we keep a set of clock reset as a label of transitions.

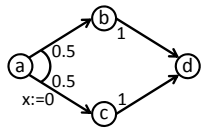


Figure 6: An Initial Abstract Model

3.2 Model Checking

In model checking, we apply Value Iteration[8] into the markov decision process obtained by abstraction and calculate a maximum reachability probability. Also, it decides an action to be chosen at every state as an adversary. If the obtained probability is less than p, we can terminate the CEGAR loop and conclude that the property is satisfied.

Although Value Iteration can calculate a maximum reachability probability, it cannot produce concrete paths used for the probability calculation. To obtain the concrete paths, we use an approach proposed in Paper[9] which can produce CE paths for PCTL formulas. The approach translates a probabilistic automaton into a weighted digraph. And we can obtain at most k paths by performing k-shortest paths search on the graph.

Definition 3.2 (Path on the Abstract Model)**.** A path $\hat{\omega}$ on an abstract model $\hat{MDP}_{PTA} = (\hat{S}, \hat{s}_0, \hat{Steps})$ for $PTA = (A, L, l_0, C, I, prob)$ is given as follows,

$$\hat{\omega} = \hat{s}_0 \xrightarrow{a_0, p_0, r_0} \hat{s}_1 \xrightarrow{a_1, p_1, r_1} \cdots \xrightarrow{a_{n-1}, p_{n-1}, r_{n-1}} \hat{s}_n$$

, where $\hat{s}_i \in \hat{S}$ for $0 \leq i \leq n$ and $(\hat{s}_i, a_i, p_i) \in \hat{Steps} \wedge (r_i, \hat{s}_{i+1}) \in support(p_i)$ for $0 \leq i \leq n-1$.

As defined in Def. 3.2, we associate a set r of clock reset with a path on an abstract model in order to show the difference of r over the probabilistic distribution p.

For the abstract model shown in Fig.6, Value Iteration outputs 1.0 as the probability that it reaches to the location d from the location a. On the other hand, k-shortest paths search ($k \geq 2$) detects two paths $\hat{\omega}^\alpha = a \xrightarrow{\tau, 0.5, \{\}} b \xrightarrow{\tau, 1.0, \{\}} d$ and $\hat{\omega}^\beta = a \xrightarrow{\tau, 0.5, \{x:=0\}} c \xrightarrow{\tau, 1.0, \{\}} d$, where τ represents a label for transitions with no label in the figure.

3.3 Simulation

Simulation checks whether all the paths obtained by k-shortest paths search are feasible or not on the original probabilistic timed automaton. We use the simulation algorithm proposed in Paper[6] where we use some operations of DBM (Difference Bound Matrix)[10] to obtain zones which are reachable from the initial state. If there is at least one path which is infeasible on the original PTA, we proceed to the abstraction refinement step.

Figure 7 shows the simulation results for two paths $\hat{\omega}^\alpha$ and $\hat{\omega}^\beta$. Simulation concludes that the two paths are feasible on the original PTA.

3.4 Abstraction Refinement

In this step, we refine the abstract model so that the given spurious CE also becomes infeasible on the refined abstract

Algorithm 1 BackwardSimulation(PTA, ω)

1: /* $PTA = (A, L, l_0, C, I, prob)$
$\hat{\omega} = \hat{s}_0 \overset{a_0,p_0,r_0}{\rightarrow} \hat{s}_1 \overset{a_1,p_1,r_1}{\rightarrow} \cdots \overset{a_{n-1},p_{n-1},r_{n-1}}{\rightarrow} \hat{s}_n$ */
2: $D^{\hat{\omega}}_{b,n} := I(\hat{s}_n)$
3: **for** $i := n-1$ downto 0 **do**
4: $D^{\hat{\omega}}_{b,i} := D^{\hat{\omega}}_{b,i+1}$
5: $D^{\hat{\omega}}_{b,i} := down(D^{\hat{\omega}}_{b,i})$ /* reverse the time elapse */
6: $D^{\hat{\omega}}_{b,i} := and(D^{\hat{\omega}}_{b,i}, I(\hat{s}_{i+1}))$
7: $D^{\hat{\omega}}_{b,i} := free(D^{\hat{\omega}}_{b,i}, r_i)$ /* remove all constraints on r_i */
8: $D^{\hat{\omega}}_{b,i} := and(D^{\hat{\omega}}_{b,i}, g_i)$ /* $(\hat{s}_i, a_i, g_i, p_i) \in prob$ */
9: $D^{\hat{\omega}}_{b,i} := and(D^{\hat{\omega}}_{b,i}, I(\hat{s}_i))$
10: **end for**
11: **return** $D^{\hat{\omega}}_b$

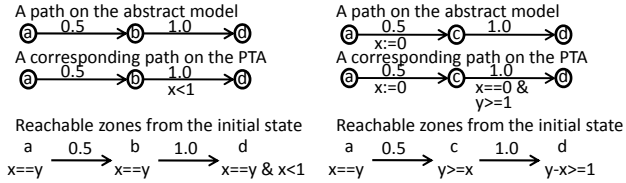


Figure 7: Simulation Results for a Set of Paths

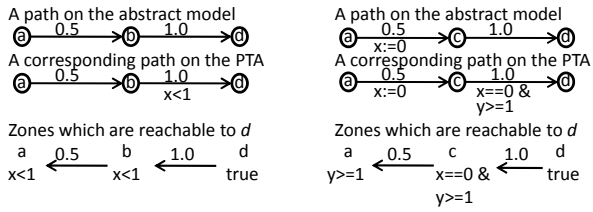


Figure 8: Results of Backward Simulation for a Set of Paths

model. We can use the algorithm proposed in Paper[6]. Since the algorithm of Paper[6] performs some operations on transitions of a timed automaton, we replace such operations by those on probability distributions of a probabilistic timed automaton.

3.5 Compatibility Checking

When all the paths obtained by k-shortest paths search are feasible and a summation of occurrence probabilities of them is greater than p, we also have to check whether all the paths are compatible or not. In this compatibility checking step, at each location of the paths, we have to obtain a condition (zone) which is reachable from the initial state and also reachable to the last state along with the path. Next, we check the compatibility of such conditions among all paths. To obtain such conditions, we have to perform both forward simulation shown in Sec. 3.3 and backward simulation for each path, and merge the results. For the result of forward simulation, we can reuse the result obtained in the Simulation step. Then we check the compatibility based on Lemma 2.2.

3.5.1 Backward Simulation

Algorithm 1 implements the backward simulation. Functions and, $free$, $down$ used in the algorithm are operation functions on a zone, and are defined in Paper[10]. Formally,

Algorithm 2 IsCompatible(PTA, $\hat{\Omega}$, D_f, D_b)

1: /* $PTA = (A, L, l_0, C, I, prob)$, $\hat{\Omega}$ is a set of abstract paths, and D_f and D_b are sets of forward and backward simulation results for each path $\hat{\omega} \in \hat{\Omega}$,respectively. */
2: **return** $CompatibleCheck(PTA, \hat{\Omega}, D_f, D_b, 0)$

Algorithm 3 CompatibleCheck(PTA, $\hat{\Omega}$, D_f, D_b, i)

1: $D' := true$
2: **foreach** $\hat{\omega} \in \hat{\Omega}$ such that $length(\hat{\omega}) \geq i$ **do**
3: $D^{\hat{\omega}}_{c,i} := D^{\hat{\omega}}_{f,i} \cap D^{\hat{\omega}}_{b,i}$
4: $D' := D' \cap D^{\hat{\omega}}_{c,i}$
5: **if** $D' = \emptyset$ **then**
6: **return** $false$
7: **end if**
8: **end for**
9: $S^{\hat{\Omega}}_{i+1} := SplitPathSet(\hat{\Omega}, i+1)$
10: /* split $\hat{\Omega}$ into a set of its subsets without overlap with respect to the i+1-th location and clock reset for each path in $\hat{\Omega}$ */
11: **foreach** $\hat{\Omega}' \in S^{\hat{\Omega}}_{i+1}$ such that $|\hat{\Omega}'| \geq 2$ **do**
12: **if** $CompatibleCheck(PTA, \hat{\Omega}', D, i+1) = false$ **then**
13: **return** $false$
14: **end if**
15: **end for**
16: **return** $true$

for a zone D, a constraint c, and a set r of clock reset, those functions are defined as follows; $and(D, c) = \{u \mid u \in D \wedge u \in c\}$, $free(D, r) = \{u \mid r(u) \in D\}$, and $down(D) = \{u \mid u + d \in D \wedge d \in \mathbb{R}_{\geq 0}\}$

Figure 8 shows results of backward simulation for two paths $\hat{\omega}^{\alpha}$ and $\hat{\omega}^{\beta}$ detected in Sec. 3.2.

3.5.2 Determination of Compatibility

In this step, we check compatibility of the set $\hat{\Omega}$ of paths on the abstract model using the required conditions obtained by both of forward and backward simulation. Algorithm 2 checks the compatibility of $\hat{\Omega}$ using the Algorithm 3.

Algorithm 3 first checks whether the required conditions of the i-th locations for each path are compatible or not ($l2$-$l8$) using the results of forward and backward simulation. Next, the algorithm divides $\hat{\Omega}$ into some subsets of it based on the (i+1)-th locations and the set of clock reset for each path ($l9$). Then, it checks the compatibility for the following sequences of paths by applying the algorithm into the divided subsets recursively ($l11$-$l15$). Although the predicate $isCompatible$

Algorithm 4 SplitPathSet($\hat{\Omega}$, i)

1: $S := \emptyset$
2: **foreach** $\hat{\omega} \in \hat{\Omega}$ **do**
3: /* $\hat{\omega} = \hat{s}_0 \overset{a_0,p_0,r_0}{\rightarrow} \hat{s}_1 \overset{a_1,p_1,r_1}{\rightarrow} \cdots \overset{a_{n-1},p_{n-1},r_{n-1}}{\rightarrow} \hat{s}_n$ */
4: **if** $\hat{\Omega}_{r_{i-1},\hat{s}_i} \notin S$ **then**
5: $\hat{\Omega}_{r_{i-1},\hat{s}_i} := \{\hat{\omega}\}$
6: $S := S \cup \hat{\Omega}_{r_{i-1},\hat{s}_i}$
7: **else**
8: $\hat{\Omega}_{r_{i-1},\hat{s}_i} := \hat{\Omega}_{r_{i-1},\hat{s}_i} \cup \{\hat{\omega}\}$
9: **end if**
10: **end for**
11: **return** S

Algorithm 5 TransformPTA(PTA, D_c, $\hat{\Omega}$, i)

1: $D_{complement}$:= true
2: **foreach** $\hat{\omega} \in \hat{\Omega}$ **do**
3: $L_{dup} := DuplicateLocation(PTA, \hat{\omega}, D^{\hat{\omega}}_{c,i}, i)$
4: $L := L \cup L_{dup}$
5: $prob_{dup} := DuplicateDistribution(PTA, \hat{\omega}, L_{dup}, i)$
6: $prob := prob \cup prob_{dup}$
7: $D_{complement} := D_{complement} \cap D^{\omega}_{c,i}$
8: **end for**
9: $L_{dup} := DuplicateLocation(PTA, \hat{\omega}, D_{complement}, i)$
10: $L := L \cup L_{dup}$
11: $prob_{dup} := DuplicateDistribution(PTA, \hat{\omega}, L_{dup}, i)$
12: $prob := prob \cup prob_{dup}$
13: $prob := RemoveDistribution(PTA, \hat{s}_i, p_i)$
14: /* for all path $\hat{\omega} \in \hat{\Omega}$, the i-th state $\hat{s}_i$ and i-th probability distribution is p_i */
15: **return** PTA

Algorithm 6 DuplicateLocation($PTA, \hat{\omega}, D, i$)

1: /* $PTA = (A, L, l_0, C, I, prob)$
$\hat{\omega} = \hat{s}_0 \overset{a_0,p_0,r_0}{\rightarrow} \hat{s}_1 \overset{a_1,p_1,r_1}{\rightarrow} \cdots \overset{a_{n-1},p_{n-1},r_{n-1}}{\rightarrow} \hat{s}_n$ */
2: $L_{dup} := \emptyset$
3: **foreach** $(l, r) \in L \times 2^C$ such that $p_i(l, r) > 0$ **do**
4: $(l, D) := Succ((\hat{s}_i, D), e)$
5: /* $succ$ returns a successor state set through a given edge e, and $e = (\hat{s}_i, a_i, g, p_i, r, l)$ */
6: $l_{dup} := newLocation()$
7: $I(l_{dup}) := D$
8: $L_{dup} = l_{dup}$
9: **end for**
10: **return** L_{dup}

in the Lemma 2.2 checks the compatibility for each subset of Ω, the algorithm omit redundant checks by dividing Ω based on the branches of the paths.

For the path $\hat{\omega}^\alpha$ in Sec. 3.2, zones at a which is reachable from initial state and which can move to d are given as $D^{\hat{\omega}^\alpha}_{f,0} = (x == y)$, and $D^{\hat{\omega}^\alpha}_{b,0} = (x < 1)$, respectively. Also, a zone of the product of them is given as $D^{\hat{\omega}^\alpha}_{c,0} = (x == y \wedge x < 1)$. Similarly, for the path $\hat{\omega}^\beta$, the product zone is given as $D^{\hat{\omega}^\beta}_{c,0} = (x == y \wedge y > 1)$. Since $D^{\hat{\omega}^\alpha}_{c,0}$ and $D^{\hat{\omega}^\beta}_{c,0}$ contradict each other, we can conclude that the paths $\hat{\omega}^\alpha$ and $\hat{\omega}^\beta$ are incompatible each other.

3.6 Model Transformation

When the compatibility check procedure decides a given set $\hat{\Omega}$ of paths is incompatible at i-th location, our proposed algorithm resolves the incompatibility by refining behaviors from the i-th location. Our algorithm uses $D^{\hat{\omega}}_c$ which is a product of results of forward and backward simulation for a path $\hat{\omega} \in \hat{\Omega}$. It duplicates locations which are reachable from the zone $D^{\hat{\omega}}_{c,i}$ by an action associated with the i-th distribution p_i. Also it constructs transition relations so that the transformation becomes equivalent transformation. For example, transition relations from a duplicated location are duplicated if the relations are executable from the invariant associated with the duplicated location.

Algorithm 5 transforms a given PTA with considering its

Algorithm 7 DuplicateDistribution($PTA, \hat{\omega}, L_{dup}, i$)

1: /* $PTA = (A, L, l_0, C, I, prob)$
$\hat{\omega} = \hat{s}_0 \overset{a_0,p_0,r_0}{\rightarrow} \hat{s}_1 \overset{a_1,p_1,r_1}{\rightarrow} \cdots \overset{a_{n-1},p_{n-1},r_{n-1}}{\rightarrow} \hat{s}_n$ */
2: $prob_{dup} := \emptyset$
3: $p_{dup} := newDistribution()$
4: /* generate a new distribution over $L \times 2^C$ */
5: **foreach** $(l, r) \in L \times 2^C$ **do**
6: $p_{dup}(l_{dup}, r) := p_i(l, r)$
7: /* l_{dup} is a duplicate location of l generated by DuplicateLocation algorithm */
8: **end for**
9: $prob_{dup} := Prob_{dup} \cup \{(\hat{s}_i, a_i, g, p_{dup})\}$
10: /* $(\hat{s}_i, a_i, g, p_i) \in prob$ */
11: **foreach** $l_{dup} \in L_{dup}$ **do**
12: $prob_{dup} := Prob_{dup} \cup$
13: $DuplicateDistFromDupLoc(PTA, l_{dup})$
14: **end for**
15: **return** p_{dup}

Algorithm 8 DuplicateDistFromDupLoc(PTA, l_{dup})

1: /* $PTA = (A, L, l_0, C, I, prob)$, and let l be an original location of l_{dup} */
2: $prob_{dup} := \emptyset$
3: **foreach** $(l, a, g, p) \in Prob$ **do**
4: f_{dup} := true, $p_{dup} := newDistribution()$
5: **foreach** $(l', r) \in L \times 2^C$ **do**
6: **if** $Succ((l, I(l_{dup})), e) \neq \emptyset$ **then**
7: /* $e = (l, a, g, p, r, l')$ */
8: $p_{dup}(l', r) = p(l, r)$
9: **else**
10: f_{dup} := false
11: **break**
12: **end if**
13: **end for**
14: **if** f_{dup} **then**
15: /* duplicate the distribution if it is executable from the duplicate location */
16: $Prob_{dup} := Prob_{dup} \cup \{(l, a, g, p_{dup})\}$
17: **end if**
18: **end for**

compatibility. The algorithm calls $DuplicateLocation$ (Algorithm 6) which duplicates locations, $DuplicateDistribution$ (Algorithm 7) which duplicates probabilistic transitions, and $RemoveDistribution$ (Algorithm 9) which removes probabilistic transitions. The procedure $Succ$ in Algorithms 6 and 8 calculates a successor state set from a given state set S through a given edge $e = (l, a, g, p, r, l')$, i.e. $Succ(S, e) = \{(l', r(\nu) + d) \mid (l, \nu) \in S \wedge g(\nu) \wedge I(l')(r(\nu)) \wedge \forall d' \leq d.I(l')(r(\nu) + d')\}$

Figure 9 shows the transformed PTA by applying the model transformation procedure for the paths $\hat{\omega}^\alpha$ and $\hat{\omega}^\beta$. The loca-

Algorithm 9 RemoveDistribution(PTA, l, p)

1: /* $PTA = (A, L, l_0, C, I, prob)$, and let l be an original location of l_{dup} */
2: **foreach** (l, a, g, p) **do**
3: $prob := prob \setminus \{(l, a, g, p)\}$
4: **end for**
5: **return** $prob$

Table 1: Experimental Result

		Digital Clocks[3]				Proposed Approach				
$D(\mu s)$	p	$Result$	$Time(s)$	$State$	$MEM(MB)$	$Result$	$Time(s)$	$Loop$	$State$	$Heap(MB)$
5	1.09×10^{-1}	false	20.90	297,232	10.2	false	4.19	10	37	8.0
	3.28×10^{-1}	true	20.89	297,232	10.2	true	3.60	9	36	8.0
10	1.26×10^{-2}	false	54.80	685,232	21.7	false	8.16	19	134	8.0
	3.79×10^{-2}	true	54.82	685,232	21.7	true	6.57	15	115	8.0
20	1.85×10^{-4}	false	176.93	1,461,232	41.0	false	1186.08	47	477	64.0
	5.56×10^{-4}	true	177.46	1,461,232	41.0	true	31.32	32	435	8.0

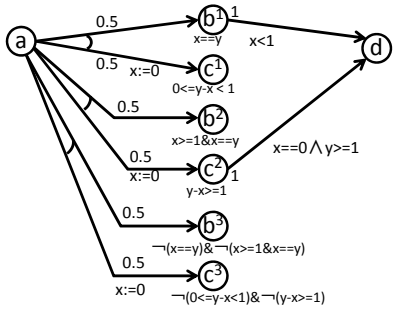


Figure 9: A Transformed PTA

tions b^1 and c^1 are duplicated locations based on the path $\hat{\omega}^{\alpha}$ and the zone $D^{\hat{\omega}^{\beta}}_{c,0} = (x == y \wedge x < 1)$ on the location a. We associate invariants to b^1 and c^1 based on zones which are reachable from $D^{\hat{\omega}^{\beta}}_{c,0}$ through transitions from a to b, and from a to c, respectively. Also, we duplicate a transition from b to d as the transition from b^1 to d because the transition is feasible from the invariant of b^1. On the other hand, we do not duplicate a transition from c to d because the transition is not feasible from the invariant of c^1. Similarly, locations b^2 and c^2 are duplicated locations based on the path $\hat{\omega}^{\beta}$ and the zone $D^{\hat{\omega}^{\beta}}_{c,0}$. Locations b^3 and c^3 are generated as complements of the invariant associated with each duplicated location in order to preserve the equivalence.

By transforming the original PTA in such a way, if we remove all clock constraints from the model in Fig.9, Value Iteration on its abstract model outputs 0.5 as the maximum probability.

4 Experiments

We have implemented a prototype of our proposed approach with Java, and performed some experiments. Though the prototype can check the compatibility of a given set of paths, currently it cannot deal with the model transformation.

The prototype performs k-shortest paths search and simulation concurrently in order to reduce execution time. By implementing the algorithms concurrently, we have not to wait until all of k paths are detected, i.e. if a path is detected by the k-shortest paths search algorithm, we can immediately apply simulation and (if needed) abstraction refinement procedures.

Also, our prototype continues the k-shortest search algorithm when a spurious CE is detected and the refinement algorithm is applied. If other paths which do not overlap with the previous spurious CEs, are detected, we can apply simulation and refinement algorithms to it again. This helps us reduce the number of CEGAR loop.

4.1 Goals of the Experiments

In this experiment, we evaluated the performance of our proposed approach with regard to execution time, memory consumption, and qualities of obtained results. As a target for comparison, we chose the approach of Digital Clocks[3] where they approximate clock evaluations of a PTA by integer values.

4.2 Example

We used a case study of the FireWire Root Contention Protocol[11] as an example for this experiment. This case study concerns the Tree Identify Protocol of the IEEE 1394 High Performance Serial Bus (called "FireWire") which takes place when a node is added or removed from the network. In the experiment, we checked the probability that a leader is not selected within a given deadline. The probabilistic timed automaton for the example is composed of two clock variables, 11 locations, and 24 transitions.

4.3 Procedure of the Experiments

In this experiment, we checked the property that "the probability that a leader cannot be elected within a given $deadline$ is less than p." We considered three scenarios where the parameter $deadline$ is 5, 10, 20 μs, respectively. Also, for each scenario, we conducted two experiments where the value of p is 1.5 times as an approximate value of the maximum probability obtained by the Digital Clocks approach[3] and a half of it, respectively. In the proposed approach, we searched at most 5000 paths by letting the parameter k of the k-shortest paths search algorithm be 5000. For evaluation of existing approach, we used the probabilistic model checker PRISM[12].

The experiments were performed under Intel Core2 Duo 2.33 GHz, 2GB RAM, and Fedora 12 (64bit).

4.4 Results of the Experiments

The results are shown in Table 1. The column of D means the value of $deadline$. For each approach, columns of $Results$, $Time$, and $States$ show the results of model checking, execution time of whole process, and the number of states constructed, respectively. The column MEM in the columns of the Digital Clocks shows the memory consumption of PRISM. The columns $Loop$ and $Heap$ in the columns of the proposed approach show the number of CEGAR loops executed and the

Table 2: Analysis of Counter Example Paths

$D(\mu s)$	p	$Path$	$Probability$	$CC(ms)$
5	1.0938×10^{-1}	7	1.2500×10^{-1}	0.7
10	1.2635×10^{-2}	43	1.2695×10^{-2}	5.9
20	1.8500×10^{-4}	2534	1.8501×10^{-4}	296.9

maximum heap size of the Java Virtual Machine (JVM) which executes our prototype, respectively.

Table 1 shows that for all cases we can dramatically reduce the number of states and obtain correct results. Moreover, we can reduce the execution time more than 80 percent except for the case when $deadline = 20\mu s$ and $p = 1.85 \times 10^{-4}$. In this case, however, the execution time drastically increases.

Table 2 shows the results of analysis of CE paths obtained when the results of model checking are false. The columns of $Path$, $Probability$ and CC show the number of CE paths, the summation of occurrence probability of them, and execution time for compatibility checking, respectively. For this example, the obtained sets of CE paths are compatible in every case.

4.5 Discussion

From the results shown in Table 1, we can see that our proposed approach is efficient with regard to both execution time and the number of states. Especially, the number of states decrease dramatically. The execution time is also decreased even though we perform model checking several times shown in the column of $Loop$.

On the other hand, in the case when $deadline = 20\mu s$ and $p = 1.85 \times 10^{-4}$, the execution time increases drastically. We think that as shown in Table 2 we have to search 2534 paths and this causes the increase of execution time especially for k-shortest paths search. A more detailed analysis shows that the execution time for k-shortest paths search accounts for 1123 seconds of total execution time of 1186 seconds. Also, the results shows that the JVM needs 64MB as its heap size in this case. This is because compatibility checking for 2534 of paths needs a large amount of the memory. From the results, we have to resolve a problem of the scalability when the number of candidate paths for a CE becomes large.

5 Conclusion

This paper proposed the abstraction refinement technique for a probabilistic timed automaton by extending the existing abstraction refinement technique for a timed automaton.

Future work includes completion of implementation. General DBM does not support not operator[13]; so we have to investigate efficient algorithms for the not operator.

Acknowledgments

This work is being conducted as a part of Stage Project, the Development of Next Generation IT Infrastructure, supported by Ministry of Education, Culture, Sports, Science and Technology, as well as Grant-in-Aid for Scientific Research C(21500036), as well as grant from The Telecommunications Advancement Foundation.

REFERENCES

[1] E. M. Clarke, O. Grumberg, and D. A. Peled. *Model checking*. MIT Press, 1999.

[2] M. Kwiatkowska, G. Norman, J. Sproston, and F. Wang: "Symbolic model checking for probabilistic timed automata," *Information and Computation*, Vol. 205, No. 7, pp. 1027–1077, 2007.

[3] M. Kwiatkowska, G. Norman, and J. Sproston: "Performance analysis of probabilistic timed automata using digital clocks," *Int. Journal on Formal Methods in System Design*, Vol. 29, No. 1, pp. 33–78, 2006.

[4] M. Kwiatkowska, G. Norman, and D. Parker, "Stochastic Games for Verification of Probabilistic Timed Automata,". In *Proc. of the 7th Int. Conf. on Formal Modeling and Analysis of Timed Systems (FORMATS'09)*, Vol. 5813 of *LNCS*, pp. 212–227, 2009.

[5] E. M. Clarke, O. Grumberg, S. Jha, Y. Lu, and V. Helmut: "Counterexample-guided abstraction refinement for symbolic model checking," *Journal of the ACM*, Vol. 50, No. 5, pp. 752–794, 2003.

[6] T. Nagaoka, K. Okano, and S. Kusumoto: "An Abstraction Refinement Technique for Timed Automata Based on Counterexample-Guided Abstraction Refinement Loop," *IEICE Transactions on Information and Systems*, Vol. E93-D, No. 5, pp. 994–1005, 2010.

[7] C. Derman. *Finite-State Markovian Decision Processes*. New York: Academic Press, 1970.

[8] D. P. Bertsekas. *Dynamic programming and optimal control*. Athena Scientific, 1995.

[9] H. Aljazzar and S. Leue: "Directed Explicit State-Space Search in the Generation of Counterexamples for Stochastic Model Checking," *IEEE Transactions on Software Engineering*, Vol. 36, No. 1, pp. 37–60, 2010.

[10] J. Bengtsson and W. Yi, "Timed Automata: Semantics, Algorithms and Tools,". In *Lecture Notes on Concurrency and Petri Nets*, Vol. 3098, pp. 87–124, 2004.

[11] M. Kwiatkowska, G. Norman, and J. Sproston: "Probabilistic model checking of deadline properties in the ieee1394 firewire root contention protocol," *Formal Aspects of Computing*, Vol. 14, No. 3, pp. 295–318, 2003.

[12] A. Hinton, M. Kwiatkowska, G. Norman, and D. Parker:, "PRISM: A Tool for Automatic Verification of Probabilistic Systems,". In *Proc. of the 12th Int. Conf. on Tools and Algorithms for the Construction and Analysis of Systems (TACAS'06)*, Vol. 3920, pp. 441–444, 2006.

[13] A. David, J. Hakansson, K. G. Larsen, and P. pettersson, "Model Checking Timed Automata with Priorities using DBM Subtraction,". In *Proc. of the 4th Int. Conf. on Formal Modelling and Analysis of Timed Systems*, pp. 128–142, 2006.

Modeling for the Organization by System Dynamics

Toshikatsu Asai*, Naokazu Yamaki**, Tadanori Mizuno***, and Sanshiro Sakai****

*Graduate School of Science and Technology, Shizuoka University, Japan
** Center for Information Infrastructure, Shizuoka University, Japan
***Graduate School of Science and Technology, Shizuoka University, Japan
****Faculty of Informatics, Shizuoka University, Japan

Abstract - The organizational behavior of enterprises is changed dynamically by various factors, such as the organizational structure, the sense of values that employees share, the leadership style, and the environment around the organization. We analyze the organizational behavior by using system dynamics. In this study, the constituent member's ways of thinking is investigated as an important factor that influences corporate activity. Constituent members are divided roughly into two types: (i) those who accept the layered structure of a corporate organization and demonstrate specific abilities under the instruction of management, and (ii) those who perform best by using their independent judgment, under the direction of management. The relationship between the approach of organizational innovation and the organizational power improvement of the enterprise has been analyzed by considering the enterprise to be a type of ecosystem consisting of the constituent members.

Keywords: System dynamics, Awareness of hierarchical order, Leadership value

1 INTRODUCTION

In this study, we recognize an organization as a system and clarify some of the transformations/movements of an organization by using a dynamic model. As the organizational innovation shifts from the sense of values that the organization already has to another sense of values, we can consider the organization as an ecosystem model consisting of members accepting new ideas and those following existing ideas. As a result, it is the purpose of this study to clarify the requirements for organization transformation, the influence of talent composition, and their relation to productivity.

2 PRIOR RESEARCH AND PROBLEM DEFINITION

2.1 System dynamics

Meadows et al. reported that the human population and economic growth in the 21st century may reach a limit because of restrictions related to limits on the earth's resources and to pollution, as identified in their 1972 Club of Rome report *The Limits to Growth* [1]. Since that time, system dynamics has become widely recognized and has been applied to many areas besides those directly related to the environment/ecosystems. These include product/technology development, the production schedule, corporate strategy and organization within a company, in addition to international relations, the public policy/social, economy/industry and energy/resources problems, which are addressed by hypothesis construction in the natural and social sciences. System dynamics can include, as part of the learning process, people's thoughts, communication, and how these can be applied to problems of the organization.

2.2 Defining the problem

The progress of informationization has increased the speed of change in the enterprise environment. The shift from the conventional pyramid organization, consisting of hierarchies, to a flatter network organization has been discussed with respect to the transformation of the organization that accompanies it [2]. For example, Kotter insisted that not only the management abilities to plan, assign resources and control, but also the leadership to present a vision, motivate members and enlighten is necessary because many companies have a management surplus [3]. Malone proposed a transformation from a management model of "command and control", which assumes centralization, to a new decentralized model: "coordinate and cultivate" [4]. In addition, Malone assumed that communications costs decide the concentration or the decentralization in the relationship between the information technology and decentralization. At first, it is decentralization in the state that each is not connected. Afterwards, it is assumed that it is centralized when they are

connected, the communications costs fall finally, and it faces decentralization. [5].

In this study, we focus on the sense of values of the constituent members and assume that each member either responds better to a pyramid organization (defined as a member with an X-type talent), responds better to a flat organization (a Y-type talent) or is intermediate, being able to shift between the two (a Z-type talent). We investigate the influence of the proportions of these three talents among the constituent members on organization transformation.

3 THE BASIC IDEA BEHIND THE MODEL

The model consists of an employee awareness model, a corporate activity model and a company environment model. (See Figure 1.)

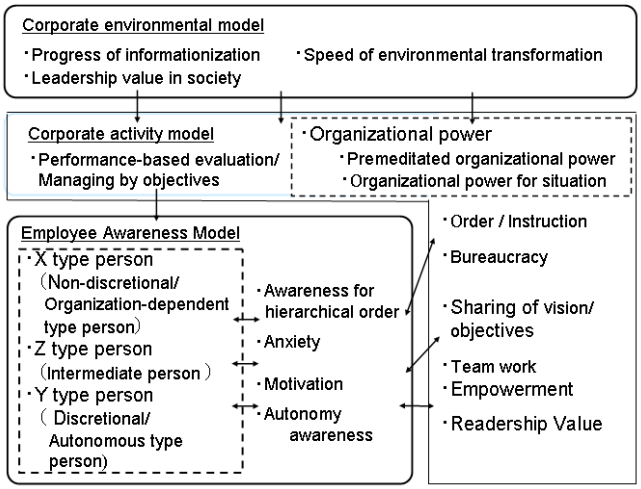


Figure 1. Basic element chart of model

3.1 Employee awareness model

(1) X-type members and Y-type members

X-type members (i.e., members with the X-type talent) prefer order in the organization. These members value the hierarchical relationship of official posts and ranks to clarify the chain of command and to advance every task according to the schedule. Execution of duties is managed and controlled by instructions to realize a plan. X-type superiors are *non-discretional* and X-type subordinates are *non-autonomous*.

In contrast, Y-type members flexibly respond to change. They also work outside the bounds of the organization and official job positions. These members think about their individual growth and the growth of the organization, aim at sharing their personal visions and objectives, and take the lead based on their degree of empowerment. Y-type superiors are *discretional* and Y-type subordinates are *autonomous*.

(2) Awareness of hierarchical order and awareness of autonomy

The basis of management by an X-type superior is to demand obedience to the superior's instructions; subordinates are expected to steadily accomplish their business by following these instructions. The way of thinking of a constituent member who supports the sense of values of the pyramid organization is called *awareness of hierarchical order*. Awareness of hierarchical order is the outlook of organizations that think of each work role as limited, focusing on the levels of the organization and assuming that all members are non-discretional and non-autonomous. When awareness of hierarchical order is controlled, *awareness of autonomy* is increased. If awareness of autonomy grows, empowerment becomes possible, and a decentralized organization can be realized.

3.2 Corporate activity model

We examine the following four factors that influence the organization movement and transformation.

(1) Strength of a management-by-objectives policy

To establish an organization that can respond to intense environmental change, management by objectives (MBO) with a performance-based evaluation is introduced. When awareness of hierarchical order is dominant, personal goals of members are determined directly by the objective of the organization. On the other hand, if MBO is executed under autonomous management of individuals in an organization that does not have awareness of hierarchical order, then awareness of autonomy is dominant. Drucker, who first advocated MBO, defined it as managing by objectives and self-control [6].

(2) Strength of leadership value

Environmental changes are expected to be large; therefore, the organization needs to promptly respond to them immediately. It becomes impossible for one person or just a few members to manage everything in the organization. Therefore, a decentralized leadership was been advocated, with assumption that leadership exists in various places within the organization [7].

In this model, the leader's most important role is personnel training. Each leader trains the next leader [8]. Leadership is required for transformation of the organization. Tichy stated that leadership needs to untie individuals from past habitual practices and find a new sense of values [9]. This sense of values was called the *leadership value* at General Electric (GE), where Tichy took part in the management training program.

The performance of Y-type members depends positively on the leadership value. When the strength of the leadership value increases, consistent personnel evaluation and training can be executed that establish the leadership value as the dominant source of values in this organization. The reliability of the organizations actions increases, and the constituent member's autonomous behavior is promoted. The strength of the leadership value depends on the leadership of the top management.

(3) Strength of vision presentation

When the leadership value becomes strong, discretion is given to individuals. But autonomous individuals can lead to a fear of losing control of the team. In this case, an organizational vision indicating the overall direction of the organization must be presented to members. If the strength of this presentation is improved, the team will not become disjointed even if authority is transferred to autonomous individuals, and management will be maintained.

(4) Ecosystem consisting of three talent types

A basic part of the model is an ecosystem model in which the X-, Y- and Z-type members are composed of three-level hierarchies of executives, managers and workers, as shown in Figure 2. An X-type member is more active and has a higher productivity in an organization with a greater awareness of hierarchical order. On the other hand, a Y-type member is more active and has a higher productivity in an organization in which the leadership value has been strengthened. If the MBO policy is strengthened, the constituent member's autonomous behavior will be repressed in X-type members and promoted in Y-type members. Strengthening the presentation of a corporate vision has a similar effect.

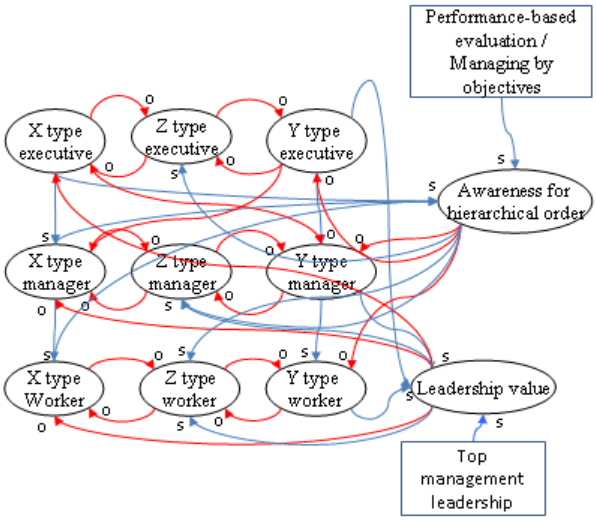


Figure 2. Employee awareness model which consists of three types of talent pattern

4 BASIC STRUCTURE AND COMPONENTS OF THE MODEL

4.1 Employee awareness model

[X-, Z- and Y-type members]

X-, Z- and Y-type members change their talent type according to the state of the awareness of hierarchical order and the leadership value. In other words, X-type executives train X-type managers, and Y-type managers train Y-type workers.

[Awareness of hierarchical order] *Ah(t)*

Awareness of hierarchical order is the primary attribute of X-type members, but Z-type members have it as well. Awareness of hierarchical order in the organization causes the following trend of change:

Y-type members ⇒ Z-type members ⇒ X-type members.

On the other hand, Y-type members tend to decrease the total awareness of hierarchical order. The awareness of hierarchical order is defined by

$$\begin{aligned} Ah(t+1) &= Ah(t) \\ &+ Ah(t)\, f_1(Mbo(t), Xe(t), Xm(t), Xw(t), Ze(t), \\ &Zm(t), Zw(t), Os(t), Ax(t), Lt(t)) \\ &- Ah(t)\, f_2(Ze(t), Zm(t), Zw(t), Ye(t), Ym(t), \\ &Yw(t), Ls(t), Lt(t), Tl(t)) \end{aligned} \quad (1),$$

where *Xe(t), Ze(t)* and *Ye(t)* are the numbers of X-, Z- and Y-type executives, respectively; *Xm(t), Zm(t)* and *Ym(t)* are the number of X-, Z- and Y-type managers, respectively; *Xw(t), Zw(t)* and *Yw(t)* are the number of X-, Z- and Y-type workers, respectively; *Mbo(t)* is performance-based evaluation and MBO (*); *Os(t)* is the organization scale coefficient; *Ax(t)* is anxiety; *Lt(t)* is top leadership (*); *Ls(t)* is the understanding leadership value in society; and *Tl(t)* is personnel evaluation and leadership training (*).

(*) Input variables

[Awareness of autonomy] *Sc(t)*

The improvement of empowerment and awareness of autonomy advances synergistically as the leadership value is strengthened, as follows:

$$\begin{aligned} Sc(t+1) &= Sc(t) + f_3(Mbo(t), Ep(t), Tl(t)) \\ &- f_4(Ax(t), Ea(t)) \end{aligned} \quad (2),$$

where *Ep(t)* is the empowerment coefficient and *Ea(t)* is the environment coefficient.

[Motivation] *Mx(t), Mz(t)* and *My(t)*

The motivation of X-type members is high in an environment with a high awareness of hierarchical order. If the leadership value is high, the motivation of Y-type members is high. Therefore, the motivation of Y-type members decreases when the motivation of X-type members increases. X-type motivation *Mx(t)* is defined by

$$Mx(t+1) = Mx(t) + Mx(t)\, f_5(IMx(t), Ax(t)) - My(t)\, f_6(IMy(t), Sc(t), Vo(t), Ax(t)) \quad (3),$$

where *IMx(t)* is the motivation coefficient of an X-type member., *IMy(t)* is the motivation coefficient of a Y-type member and *Vo(t)* is the sharing of vision and objective.

4.2 Corporate activity model

[Performance-based evaluation and MBO] *Mbo(t)*

This input variable takes on values 0 to 5, where 0 indicates a seniority system unaffected by results, and 5 indicates that performance-based evaluation and MBO are thorough, and the strategy of management has been established by balanced scorecard (BSC).

[Top management leadership] *Tl(t)*

This input variable takes on values from 0 to 5. The value 0 indicates that the top management does not exercise transformational leadership, and 5 indicates that transformational leadership is exercised most strongly.

[Leadership value] *Lv(t)*

The performance of Y-type members in the organization depends on the leadership value. This is true for executes, managers and workers. Autonomous individuals are trained, and an organization that is learning and growing sustainably is established. The leadership value accumulated in the organization is defined by

$$Lv(t+1) = Lv(t) + Lv(t)\, f_7(Ls(t), Ye(t), Ym(t), Yw(t), Ze(t), Zm(t), Zw(t), Lt(t), Tl(t)) - Lv(t)\, f_8(Xe(t), Xm(t), Xw(t), Ze(t), Zm(t), Zw(t), Os(t), Ax(t), Lt(t)) \quad (4).$$

[Empowerment coefficient] *Ep(t)*

Empowerment means the organization shares the leadership value, autonomy is granted to constituent members, and the empowerment of members is attempted. However, empowerment does not advance well if autonomy is insufficient. It is necessary to construct a virtuous circle among the sharing of the leadership value, the promotion of empowerment, and individual autonomy. Empowerment is defined by

$$Ep(t+1) = Ep(t) + Ep(t)\, f_9(Sc(t), Ea(t)) - Ep(t)\, f_{10}(Ea(t)) \quad (5).$$

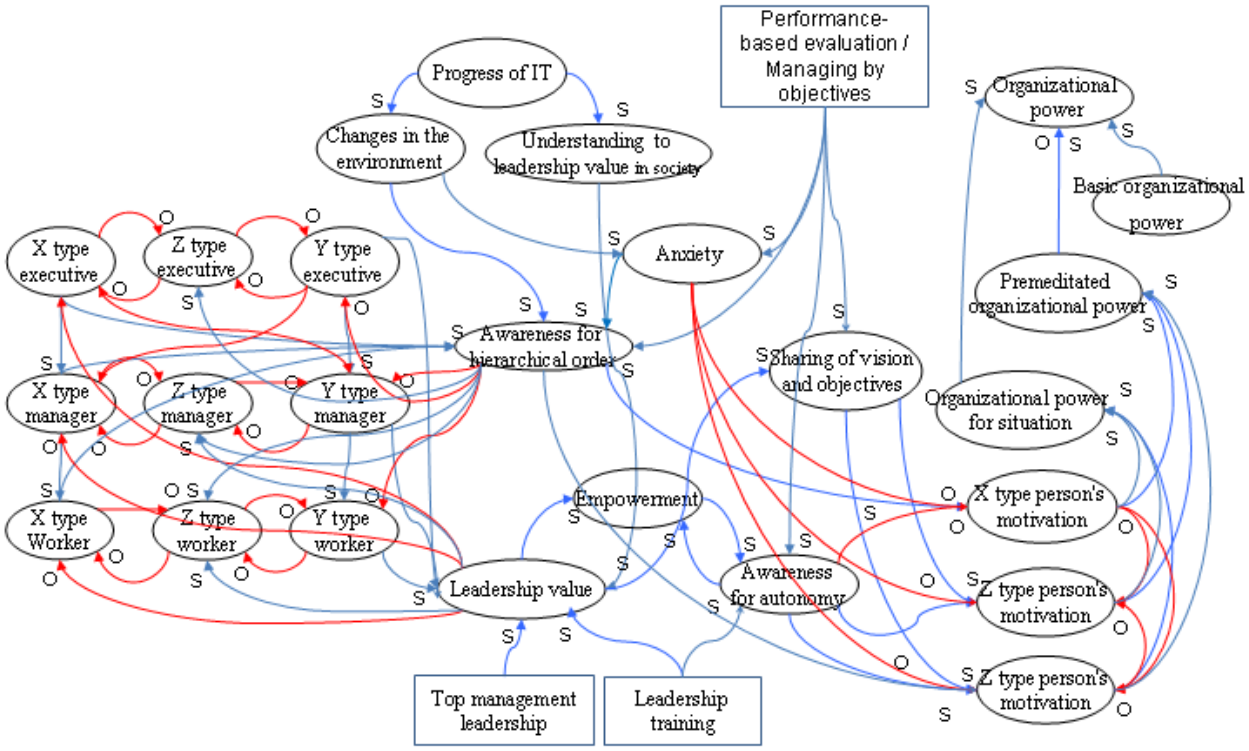


Figure 3: Information-based organization model

[Sharing of vision and objectives] *Vo(t)*

Individual participation in the decision making of the organization is attempted through discussion to promote individual autonomy by empowerment under performance-based evaluation and MBO. Commitment to the organization is sought by sharing a vision and objectives:

$$Vo(t+1) = Vo(t) + Vo(t) f_{11}(Mbo(t), Ea(t)) - Vo(t) f_{12}(Mbo(t), Lv(t), Ah(t)) \quad (6).$$

[Organizational power] *Op(t)*

This index indicates the final productivity of the enterprise but is not an individual financial index of the enterprise. In other words, this index shows the energies of the organization, and each financial index improves if this organizational power rises. The organizational strength is the total of premeditated organizational power, situational organizational power and basic organizational power:

$$Op(t) = Opb(t) + Opp(t) + Opa(t) \quad (7),$$

where *Opb(t)* is the basic organizational power,
Opp(t) is the premeditated organizational power and *Opa(t)* is the situational organizational power.

[Premeditated organizational power] *Opp*(*t*)

X-type members are characterized by working according to a plan, and *premeditated organizational power* is primarily present in these members. It is defined by

$$Opp(t) = 0.8 * Opx(t) + 0.5 * Opz(t) + 0.2 * Opy(t) \quad (8),$$

where *Opx(t), Opz(t) and Opy(t)* are the organizational power of X-, Z - and Y -type members, respectively. The organizational power according to each type is calculated by multiplying the number of each type of members by each motivation coefficient.

[Situational organizational power] *Opa(t)*

Those with Y-type talent work while staying in step with change. The situational organizational power is present primarily in Y-type members:

$$Opa(t) = 0.2 * Opx(t) + 0.5 * Opz(t) + 0.8 * Opy(t) \quad (9).$$

Basic Organizational Power

This is power specific to an organization and not related to environment transformation.

4.3 Corporate environmental model

[Understanding the leadership value in society] *Ls(t)*

Any paradigm shift of an enterprise advances step by step. The employees recognize through seminars outside the company or through the mass media that they should have a different sense of value than the current one in the organization. This is thought to be the promotion of understanding in society of the leadership value. The information-based organization model, in which the corporate activity model and the corporate environmental model are combined with the employee awareness model, is shown in Figure 3.

5 VERIFICATION OF OUTCOME BY MODEL EXECUTION

The three input variables of the model are as follows: top management leadership, personnel evaluation and leadership training, and performance-based evaluation and MBO. Each input variable takes values from 0 to 5, in which 0 indicates that the state is not executed or does not function, and 5 indicates that the state works most strongly or functions.

In this model execution, three variables are changed at the same level, and it goes in each revolution level. The case where the personnel evaluation and leadership training, and the performance-based evaluation and management by objectives corresponding to a top management leadership are implemented is assumed.

The future, to measure the effect of a personnel evaluation and the leadership training and the influences of the performance-based evaluation and management by objectives individually, these were assumed to be an input variable.

Level	0 1 2 3 4 5
Top management leadership	Pyramid organization ⟶ Network organization Business process innovation (producer based→consumer based) Management (command and control)→Leadership（coordinate and cultivate) Exclusion of awareness of hierarchical order → Establishment of leadership value
Personnel evaluation and leadership training	Personnel evaluation and training based on Leadership value From education to human development Coaching/Commitment training Business supplementary training → Business synchronization training Employee satisfaction index, 360-degree evaluation implementation/established Autonomous individuals / Empowerment
Performance-based evaluation and MBO	Clarification of vision Sharing of vision and strategy NORM management Cooperation in objective of organization and person Performance-based evaluation From the management of the result to the management of the process Balanced scorecard implementation/spread/established

Figure4. Three input variables

Figure 4 shows the measures corresponding to five from level 0 of these three input variables.

The simulation is executed under the following conditions:

- Six cases having fixed transformation levels (0 to 5)

Top management leadership, personnel evaluation and leadership training, and performance-based evaluation and management by objectives: six cases, adjusting all three input values to 0, 1, 2, 3, 4 or 5, respectively.

- Case of raising the transformation level gradually (0→5)

Top management leadership, personnel evaluation and leadership training, and performance-based evaluation and management by objectives: the case in which all variables improve over time, from 0 to 5.

- Initial proportions of the three types of personnel:

X–type members: 40%, Z-type members: 50%, and Y-type members: 10%.

- Simulation time: 3,600 steps (600 steps per year).

The case in which the three input values are all 0 is called *transformation level* 0. The change of the organizational power per person is shown in Figure 5 as an index that shows the productivity of the organization. Figure 6 shows the change in the number of the X-type, Z-type and Y-type managers in the cases of transformation levels 0, 1, 2 and 0→5.

(1) Six cases having fixed transformation levels (0 to 5)

Figure 5 shows the results for each transformation level. The organizational power of transformation level 1 is slightly lower than transformation level 0 for a long time but exceeds it in the last stage (from approximately step 3,000). The simulation results change greatly at transformation level 2. For transformation level 2, the loss due to transformation continues and the organizational power is sluggish at first, but it turns upward from approximately step 2,000 and then increases. In the case of transformation level 3, an initial decline of organizational power occurs again, but the depression period is shorter than that of transformation level 2. If the transformation level is gradually improved, as with 4 and 5, the period of the initial organizational power decrease shortens more, and the organizational power improves immediately.

Although there were no large changes in the organizational power for transformation levels 0 and 1 (Figure 5), the composition of X-, Z- and Y-type managers did change (Figure 6). The number of X-type mangers increased continuously for transformation level 0 until almost all mangers were X-type by the end. X-type managers are the most dominant in transformation level 1, too, but the three types of mangers are approximately balanced from step 900.

For transformation level 2, Y-type managers increase continuously through the simulation, and almost all managers are Y-type by the end.

(2) Case of raising the transformation level gradually (0→5)

The transformation level starts at 0 and is raised continuously to 5. As shown in Figure 6, at first the organizational power

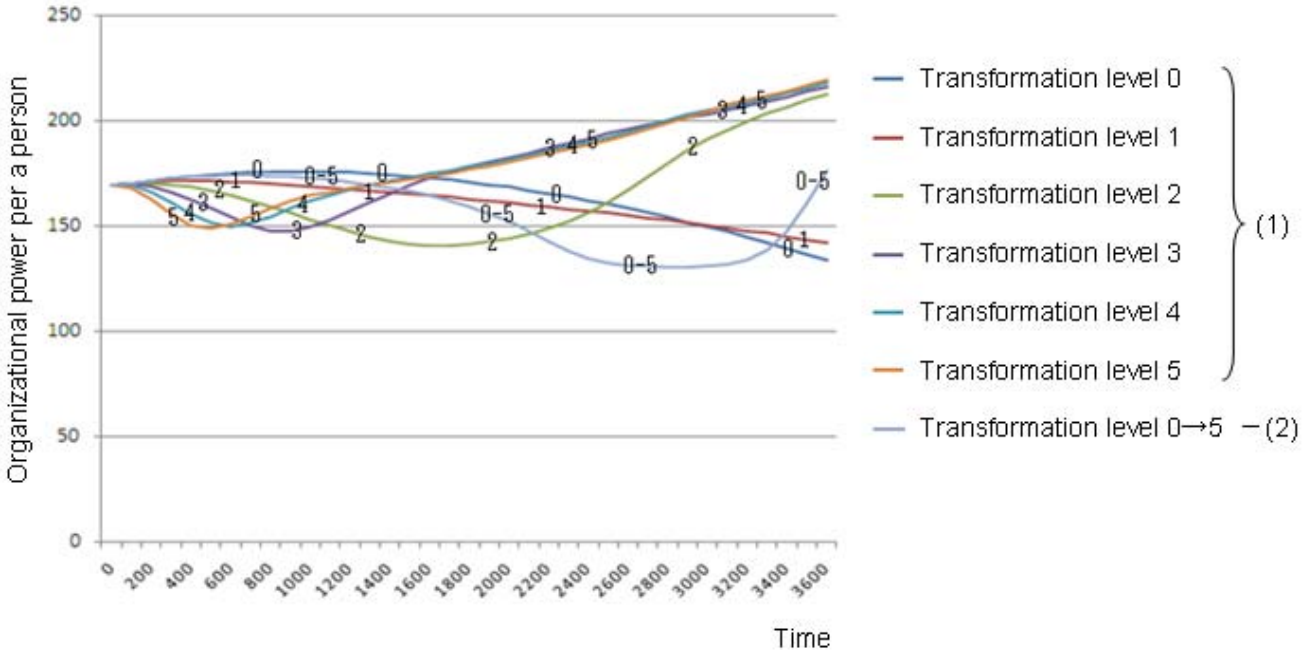


Figure 5:Transformation level and Organizational power

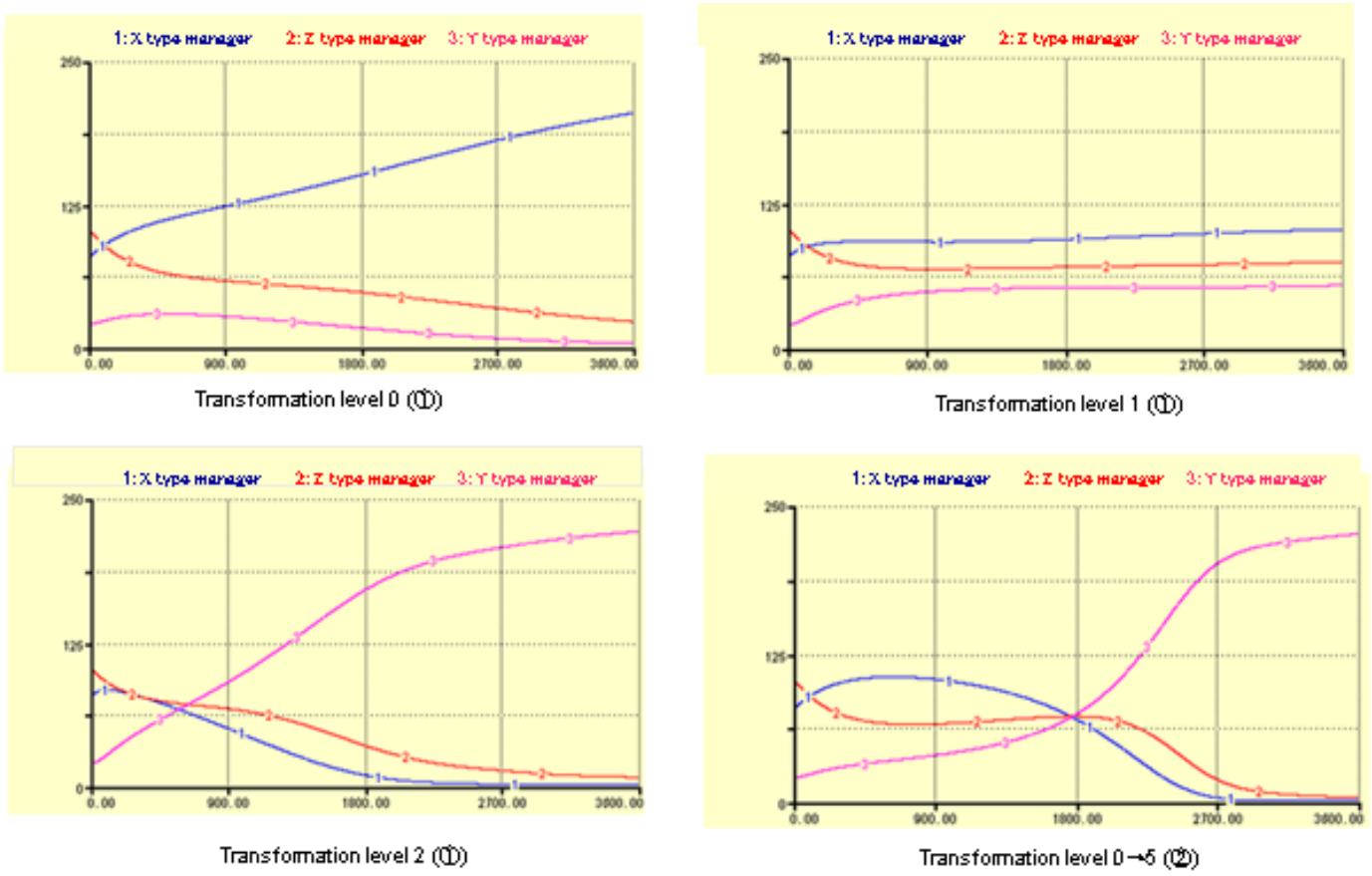


Figure 6: Transformation level and the number of managers

improves, but it decreases from approximately step 1,500. The organizational power is sluggish for a long time in the second half of the simulation, although it improves toward the end. In this case, the transformation level increases at a rate of 1 per year..

This might be the most realistic case for an organization to improve the transformation level while gradually clarifying its vision. However, the simulation result shows that the transformation loss is very large for gradual improvements in the transformation level (Figure 5). In Figure 6, the number of X-type members increases initially and the three types are balanced at step 1,800. However, the Y-type members begin to increase after step 1,800, and the sluggish organizational power eventually rises.

6 CONCLUSION

The dynamics of organizational innovation is viewed as an ecosystem by members who accept a new sense of values and members who keep their past sense of values. The tall poppy is attached to the most responsible position in the organization according to the sense of previous values. The more past successful experiences exist, the more difficult it is to take on a new kind of thinking. This is what makes it difficult for the organization to innovate. It is our belief that the challenge that Japanese organizations are now facing is to escape from the sense of value of the pyramid organization by eliminating awareness of hierarchical order and acquiring a new sense of organizational value.

In this study, the relation between the paradigm shift and organizational strength has been clarified by simulation. Transformation seems to break through the thick wall (in a word, breaking down bureaucracy), and it was clarified that it is preferable for top management leadership to transform quickly and decisively.

However, the transformational vision to establish a new organization and people has not actually been clarified in many organizations. We conclude that the result does not appear in the productivity of the organization because of an incomplete approach in which the transformation level is low from the beginning to the present.

In the future, we will investigate actual conditions. We intend to clarify the relation between the prediction results obtained by this model and reality.

REFERENCES

[1] Meadows, D. H., Meadows, D. L., Randers, J. and Behrens III, W. W., *The Limits to Growth*, Universe Books, 1972.

[2] Argyris, C. and Schon, D. A., "*Organizational Learning: A theory of Action Perspective*", Addison-Wesley, 1978.

[3] Kotter, J. P., "*What Leaders Really Do*", Harvard Business Review May-June 1990.

[4] Malone, T. W., "*The Future of Work*", Harvard Business School Press, 2004.

[5] Wyner, Geoge M. and Malone, Thomas W. ,
"Cowboys or Commanders: Does Information Technology Lead To Decentralization?", International Conference on Information Systems、1996, pp.63-79

[6] Drucker, P. F., "*The Practice of Management*", Elsevier, 1954, pp. 105-118.

[7] Ancona, D., Malone, T. W., Orlikowski, W. J. and Senge, P. M., "In Praise of the Incomplete Leader", Harvard Business Review 2007, Feb, pp, 92-100.

[8] Tichy, N. M., "*The Leadership Engine: How Winning Companies Build Leaders at Every Level*", Harper Collins Publishers, 1997.

[9] Tichy, N. M. and Devanna, M. A., "*The Transformational Leader: The Key to Global Competitiveness*", Wiley Management Classic, 1997.

[10] Asai T., "*Transformation of People and Organization with Paradigm Shift*", Communication of Japan Industrial Management Association Vol. 17, No. 3, 2007, pp. 233-237.

Problem and solution of delay in UHF belt RFID

Masashi Torato** Toshihisa Kamemaru**
Tadanori Mizuno*** and Sanshirou Sakai****

* Mitsubishi Electric Information Systems Corporation
torato-masashi@mdis.co.jp
** Mitsubishi Electric Corporation
Kamemaru.Toshiaki@ak.MitsubishiElectric.co.jp
***Graduate School of Science and Technology, Shizuoka University, Japan
mizuno@mizulab.net
****Faculty of Informatics, Shizuoka University, Japan
sakai@inf.shizuoka.ac.jp

Abstract - We show the problem of the UHF RFID system caused by the LBT that readers are influenced by other readers and do not keep real time capability when readers work simultaneously. We also show the solution for this problem and keep the real time capability by changing the session ID and suspending the transmission of the non-real time readers between rounds.

Keywords: Radio Frequency Identification, anti collision, realtime system, non-realtime system, time sharing

1 INTRODUCTION

RFID (Radio Frequency Identification) is important hardware for the ubiquitous society in which everything is connected easily via "Everyone anytime and anywhere anything" network. There are some RFID technologies and the UHF (Ultra High Frequency, 860-960MHz) RFID is one of the key technologies because the access distance (-3m) is longer compared with other technologies and it is expected to apply logistic management system and so on.

In Japan, high power (1W output) UHF RFID can be used from April 2005, low power (10mW output) can be used from January 2006. On the same time, career sense mechanism called LBT (Listen Before Talk) is introduced to prevent interference between RFID readers. In the Japanese LBT, 9 and 14 channels are defined for high power and low power respectively and if the all channels are used, then the reader must wait until other reader releases a channel.

It is easy to imagine that some readers must wait and cause read misses by the LBT when more than 9 readers work at same time [1],[2],[3]. It means that the real time RFID readers such as gate monitor loose the RFID tag when other non-real time readers such as stock control work simultaneously. In this paper, we discuss about the interference between real-time readers and non-real time [4],[5],[6], and propose the solution.

2 INTERFERENCE IN UHF RFID

2.1 Discussion about interference

One of the key features of UHF RFID is the communication distance. The tag without the battery (passive tag) can be accessed from 3m distance. It means that the response (reflected wave) from a tag is extremely small and the transmission wave of very far reader may cause the interference. So it is necessary to prevent the interference and the technologies for this purpose are called anti-collision [7],[8][9],[10].

2.2 LBT and the problem

The LBT mechanism is introduced from January 2006, so reader-reader interference problem was solved. In the LBT for the high power RFID, we can use 9 channels and sense career 5 to 10 milliseconds before transmission, then start transmission maximum 4 seconds, sleep 50 milliseconds before next transmission.

It is possible to prevent interference between readers by sensing career and waiting for a channel release, however, the reader must wait when all channels are used. As a result, delay may be generated, and the possibility that it becomes impossible to read in real time comes out. It is important for the real time RFID system to evaluate this delay and how to solve.

3 MODELING OF SIMULATION

3.1 Operation overview of Gen2

The UHF RFID technology called Class-1 Generation-2 (alias Gen2)[11] settled on by EPC global is now a main stream, though there is various kinds of UHF RFID. Figure 1 shows the operation overview of the protocol of the Gen2 and the above-mentioned Radio regulation. Reading tag is often executed two or more times usually. It is called a cycle here. LBT of 50 milliseconds is needed between cycles.

Next, the processing that is called a round following initialization is repeated at each cycle. Initialization can be divided roughly into initialization according to the processing system of the reader and the initialization of tag (session initialization). In the round, it is composed of the round initialization and the slot. In addition, the slot can be divided into the inquiry and the reading to tag. When the corresponding tag doesn't exist, the reading time becomes unnecessary.

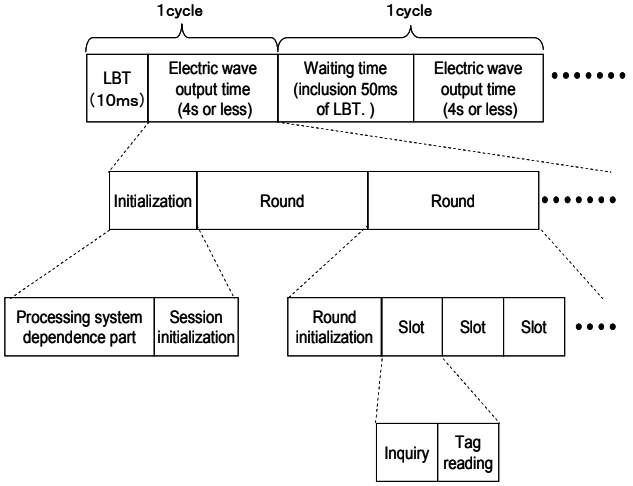


Fig1. Operation overview of Gen2

The formula that the reader puts out the electric wave is shown by the next expression.

T(cycle) = T(processing system dependence part) + T(session initialization) + the number of Round × {T(round initialization) + T(inquiry) × the number of Slot +T(tag reading) × the number of reading tag} ····· (1)

The transmission rate by one channel (200kHz) becomes 40kbps. The above-mentioned value in that case reaches the following values.

T(processing system dependence part)	about 10ms
T(session initialization)	about 2ms
T(round initialization)	about 1.5ms
T(inquiry)	about 0.8ms
T(tag reading)	about 6ms

If there is one round in which there are 16 slots:

T(cycle)=12ms+(1.5ms+0.8ms×16)+the number of tag × 6ms
=26.3ms + the number of tag×6ms ····· (2)

Multi-cycles are necessary for practical use. In that case, the number of reading times is given by the following expressions.

N(the number that can be read) = T(can be read)/(T(cycle) + T(LBT)) ···· (3)

T(LBT): Ten milliseconds, and 50 milliseconds first time at the following reading.

3.2 Model of gate monitor (realtime processing)

In UHF RFID, its accessible distance is 2-3m. The gate monitor should complete reading while the person and the thing move such a distance. For instance, the passing speed of the person is about 5km per hour, and it is assumed about 10km per hour in the belt conveyer. It becomes per second with 1.4m and 2.8m if it mends respectively. Therefore, time that T can be read becomes about one second. Having surely to read tag in this time limit, and excluding the influence of the delay by LBT become indispensable.

On the other hand, the number of tags can be assumed several pieces (a person and several things). We assume 50 milliseconds as T(cycle) in the following simulation. 50 milliseconds mean that there are 4 tags and read them in one round which has 16 slots.

3.3 Model of stock control (non-real time processing)

When the stock is taken an inventory, there are a lot of tags and they can be read at once. The slots per rounds are assumed 16, and the typical T(time of read per cycles) are followings.

T(cycle)=161 ms (20 tags and 2 rounds)
T(cycle)=235 ms (30 tags and 3 rounds)

In the simulation in chapter 4, T(time of read per cycles) are changed from 100 milliseconds to 500 milliseconds..

4 SIMULATION AND PROBLEM

4.1 Simulation results of gate monitor

Traffic that passed the gate was given by the Monte Carlo method by using these parameters, and the relation between the number of gates and the amount of traffic was simulated. After an event (i.e. a human pass the gate), the next event is generated evenly within the fixed time as following.

T(mean time)×2 – T(can be read)

For instance, 900 events per hour means T(mean time) is 4 seconds, and if 1 second is assumed for T(can be read), then the next event is generated within 7 seconds. In this simulation, event distribution within 7 seconds is uniform and share 9ch by all the gating arrangements so that -74dBm at the sense level may correspond to the distance of about 7km for the output of 36dBm of the high power type.

As mentioned, LBT may cause delay for each read cycle if there are many readers. The delay is calculated in this simulation. Figure 2 shows the relation between the delay and probability distributions of the delayed cycle. Becoming the probability of 99.999% becomes 20 milliseconds at 20 gates and 170ms at 25 gates. It means there is an influence by LBT at 25 gates on these conditions.

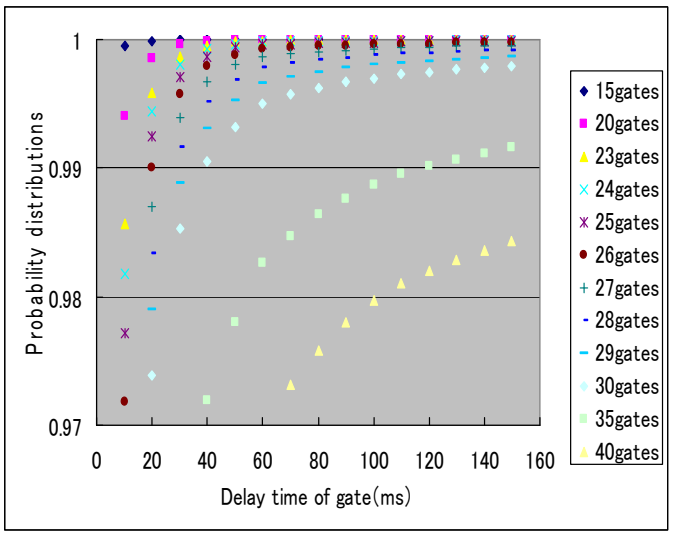


Fig2. Relation between the delay and probability distributions

Figure 3 shows the relation between the number of gates and the delay at the probability of 99.99% and 99.999%. It means that the influence can be suppressed up to 25 gates.

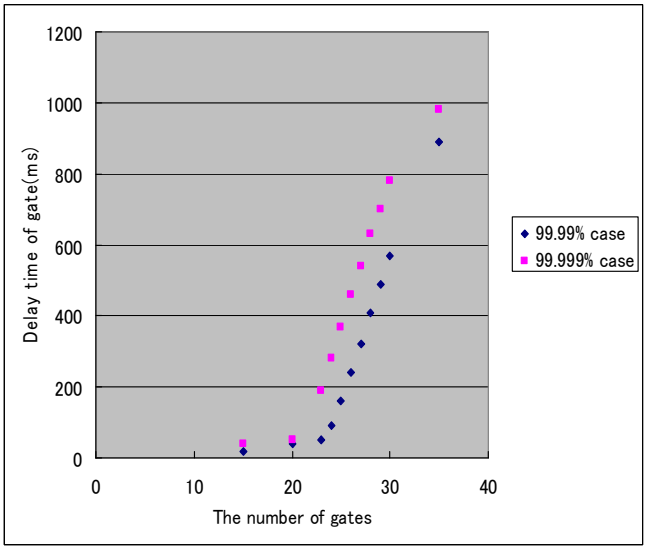


Fig3. Relation between number of gates and delay time

On this simulation, T(cycle) is 50 milliseconds, typical T(LBT) is 50 milliseconds, and T(can be read) is 1 second, so it is easy to understand that expected cycles are 10 per event. But there are some delays and the expected cycles may be decreased. Figure 4 shows the relation between the number of gates and probability at the expected cycles of 60 to 80 percent.

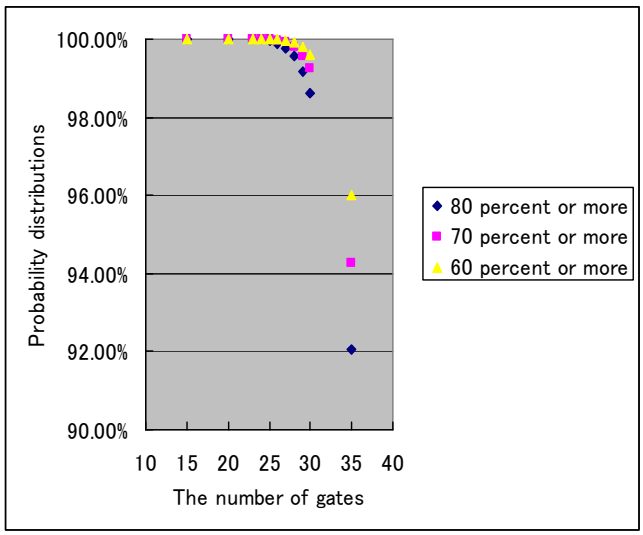


Fig4. Relation of number of gates and the probability

It also means that the boundary of the influence is about 25 gates.

4.2 Simulation results of gate monitor and stock control

Consider that there are two systems, the first one is a gate monitor system discussed above, the second one is a system that reads tags in bulk like an inventory control system, and there are at the same building. T(cycle) for the second system is a several hundreds milliseconds typically.

The simulation conditions are:

Gate monitor system (same with 4.1):
- Number of gates: 25
- T(cycle): 50 milliseconds
- T(mean time): 2 seconds
- T(can be read): 1 seconds

Stock control system:
- Number of stock controls: 5
- T(cycle) for stock control: 100～500ms
- Interval time to the next read: 500 milliseconds

Figure 5 is similar with the Figure 2 in case of 25 gates, with changing T(cycle) of stock control from 100 milliseconds to 500 seconds,

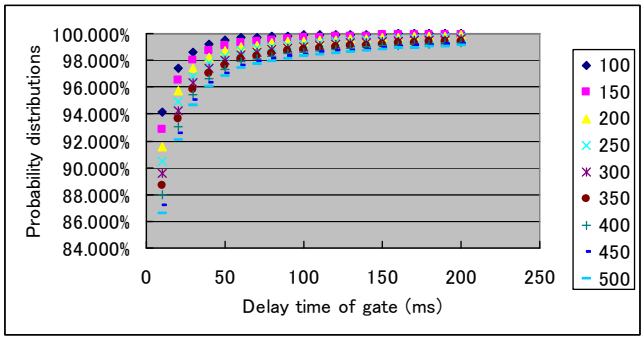


Fig5. Delay time and probability distributions

Figure 6 shows the relation between T(cycle) of stock control and the delay in the gate system at the probability of 99.99% and 99.999%.

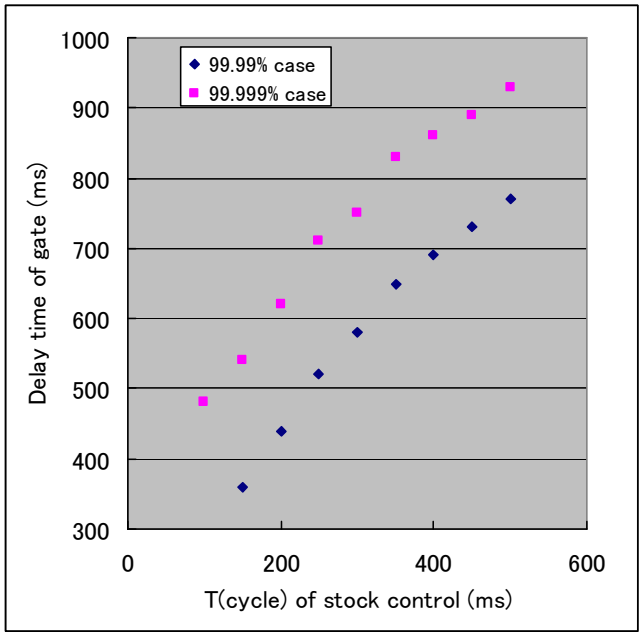


Fig6. Reading time of gate delay time and stock control

4.3 Summary of problem

When the stock control system coexists with the gate system, it is understood that the delay time is increased. It means that the number of reading cycles at the gate may decrease, and the reading accuracy at the gate where real time is necessary worsens when the reading time of the stock control becomes long.

5 SOLUTION AND VERIFICATION

As described in the section 3, a cycle consists of session initialization and one or more rounds. In the Gen2 standard, there are four sessions and some sessions can maintain the state during the fixed time even if the access by the reader stops. For instance, session S2 and S3 are able to maintain the state for two seconds or more.

We propose that either of S2 or S3 session is used on the stock control side, and other sessions are used on the gate side. In addition, we stop the reader between rounds on the stock control, and restarts after LBT (Figure 7).

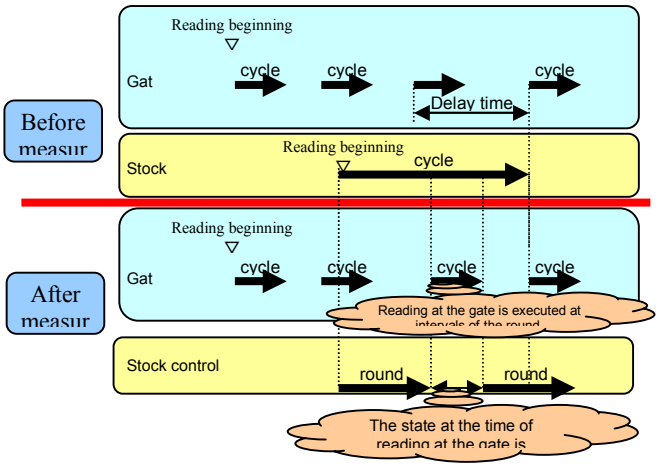


Fig7. Sequence of measures

We assume that 8 tags can be read in each round and simulate with T(round) for stock control as 75 milliseconds. Figure 8 shows the result.

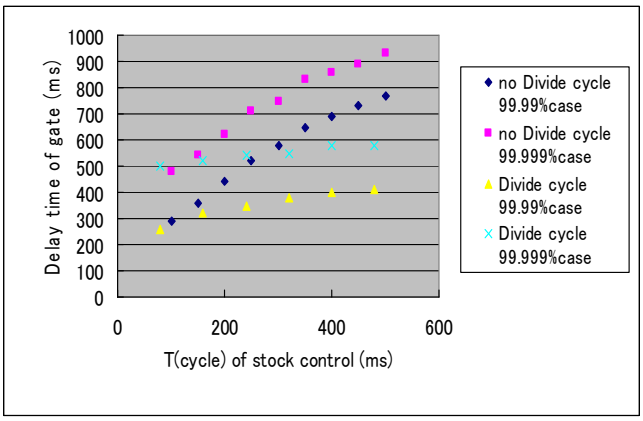


Fig.8 Delay relation between inventory system reading time and gate system

The “divided cycle” is our proposed method and Figure 8 shows the improvement by the proposed method. It shows that gate reader may be waited for 600 milliseconds by our method while it may be waited for more than 900 milliseconds by earlier method when stock control readers use channels for about 500 milliseconds.

6 SUMMARY

In this paper, we show the problem of the UHF RFID system caused by the LBT that readers are influenced by other readers and do not keep real time capability when many readers work simultaneously. We also show the solution for this problem and keep the real time capability by changing the session ID and suspending the transmission of the non-real time readers between rounds.

REFERENCES

[1] Masashi Shimizu, "The Problem of wireless telecommunications in use of ID tag", and COMPUTER & NETWORK LAN, pp.16-81, ohm company, and July 2004

[2] Yasushi Sogabe, Kamemaru Toshihisa, Akinori Taira, "Interference evasion technology between readers of UHF belt electronic tag", Monthly EMC, No.202, pp.86-93, 2005

[3] Yasushi Sogabe, Hideto Aikawa, Tsutomu Iwahashi, Toshihisa Kamemaru "Interference problem between systems in UHF belt electronic tag system",IEICE Transactions magazine, Vol.89 No.2 pp156-161,2006

[4] Akinori Taira, Toshihisa Kamemaru, Akira Kurita, Yasushi Sogabe, Fumio Ishizu, "Examination concerning UHF belt RFID interference workaround", 2005 IEICE , B-5-107, March 2005

[5] Yoshinori Tanaka, Akinao Ninomiya, Hideyuki Kobayashi,"Reader interference control method in UHF belt RFID system", 2006 IEICE ,B-5-147,March 2006

[6] Satoshi Arakawa, "Problem and measures (4) of UHF belt RFID examination of workaround of the interference between electronic tag - reader/writer device",2006 IEICE ,B-5-16,March 2006

[7] Hitoshi Mitsugi, "Decentralized interference evasion between RFID by channel selection with weight", 2006 IEICE, AS-3-12,Sept 2006

[8] Takehiro Ito,Koji Nishikawa, Hidetoshi Funakura, Satosi Arakawa, Toshihisa Kamemaru, "Proof of the interference evasion method between adjacent systems by multistep floor LBT", 2007 IEICE, B-7-126, March 2007

[9] Takehiro Ito, Koji Nishikawa, Satoshi Arakawa, Toshihisa Kamemaru, Yasushi Takahata," Problem and measures of UHF belt RFID -Method examination to making in real time-",2006 IEICE,B-5-15, March 2006

[10]Telecommunication council,Telecommunication technological subcommittee meeting,Small electric power wireless system committee report (2005 fiscal year)

[11] Specification for RFID Air Interface EPC™ Radio-Frequency Identity Protocols
Class-1 Generation-2 UHF RFID Protocol for Communications at 860 MHz − 960 MHz Version 1.0.9 Copyright notice 2004, EPCglobal Inc

Distributed System with Portable Device Based on Gesture Recognition

Osamu Yuuki, Hiroshi Mineno and Tadanori Mizuno

Graduate School of Science and Technology, Shizuoka University, Japan
osamu.yuki@canon.co.jp

Abstract – 3D imaging needs the high processing throughput. For a lot of calculation processing, we usually use big size computers. Such a computer occupies the large space in an operation room, and persons around the computer feel the oppression. By using the proposed distributed system, it is possible to handle the data between the distant places. The portable device in distributed system can perform the data processing without having many inside resources. Operators of the client feel such that one database performs all processing. The second topic concerns the control of device. Operators of the client can control the proposed portable device by gesture actions. The portable device of lap top type has not high calculation performance. In addition, because the laptop device is using such as handheld devices like keyboard, mouse, laser-tracked pen and stylus, one hand of the operator is occupied for its operating action. On the other hand, it is considered a natural communication to convey information in interactions. We demonstrate the portable device operated by gesture actions. By this method, a hand of the operator can freely be used for other action.

Keywords: distributed system, portable device, gesture recognition

1 INTRODUCTION

Three-dimensional (3D) imaging needs the high processing throughput. For example, Optoacoustic imaging in the field of medical diagnostics has to solve equations of optoacoustics wave. For a lot of calculation processing, we usually use big size computers. Such a computer occupies the large space in an operation room, and the people around the computer feel a sense of being oppressed. By using the proposed distributed system, plural people can share the information between distant places. Because it is possible to handle the data between the distant places, the portable device in distributed system can perform the data processing without having so many resources. Operators of the client feel that one database performs all processing[1]. Of course, many portable devices can exist in a network. Better and simpler software development tools along with cheaper electronics make it possible to embed web interfaces into small and inexpensive microprocessor based devices. This paper propose the concept of web embedded imaging device. We propose and design a wireless distributed data acquisition system, too. The networking system include the client of wireless-portable-device.

There have been some studies related to wireless client device. This type of system can be used as tele-medical systems. The client devices in this environment would monitor standard physiological signals: Electro-cardio-gram (ECG), Electro-encephalo-gram (EEG), oxygen saturation (SO2), breathing, temperature, [2][3] etc. The data of acquisitions would be processed by the client devices. Then the data is stored on their secondary storage until the server could download and display the data. The physician monitoring a patient, connected to the client devices, would carry around his mobile device to investigate/monitor the measurement from the patient [4]. The major requirements of the monitoring devices in such a wireless distributed data acquisition system are long battery life, lightweight, small size and big storage capacity[5][6][7].

However, these devices communicate mainly one way to the server. Usually the client devices cannot have high calculation performance. In this paper, we describe the study of a "wired or wireless distributed-data-acquisition-system based on intelligent portable device".

The second topic concerns the control of device. The operator can communicate with machines via handheld devices like keyboard, mouse, remote control, laser-tracked pen or stylus. If the laptop device is using handheld devices like keyboard, mouse, laser-tracked pen and stylus, one hand of the operator will be occupied for its operating action. The operator may try to communicate with machines via speech or activity/gesture. Because a variety of spontaneous gestures, such as finger, hand, body and head movements are used to convey information in interactions among people, among these methods, gestures can be considered a natural communication channel [8]. Operators of the client can control the proposed portable device by gesture actions. We demonstrate the portable device operated by gesture actions. By this method, a hand of the operator can freely be used for other action.

2 SYSTEM STRUCTURE

2.1 Client-Server Distributed System

We describe the structure of our proposed client-server distributed system. The portable device as a client transmits the acquired data to server. Server constructs the 3D image from the acquired data. After processing, all or parts of the data in server are transmitted to a client. Figure 1 shows the system organization. The network is able to organize multi-clients/servers by sharing the data stored in the server. A monitor of client displays the functional image by 3D. Volume rendering visualize a 3D dataset in a single image. Processing such images can be very compute intensive. And good high quality renderings need high throughput performance and much storage place. 3D texture hardware

can be addressed by three indexes (i, j, k). Index k is a depth of image. In other words, it represents time series. Three indexes present a cubic voxel data set. Such images may be stored for viewing later or displayed in real-time. While simple images can be produced rapidly, more realistic and complicated higher-resolution images can now be produced in more reasonable amounts of time. The desire to create high-quality work rather than simply wanting the same images created faster drives the need for increased computing power. The performance of the big server is typically limited by the performance of data room's cooling systems and the total electricity cost rather than by the performance of the processors. The computers, routers, power supplies and related electronics are typically mounted on racks in a server room or data center. Servers are too huge to set in an operation room, now. Recently, one of our researches has explored the feasibility of reprogramming modern video cards to do rendering in the card's hardware. The GPU has many streaming processors [9]. If GPU is used in the server, the calculation speed will become faster than processing by several multi-core-processor.

On the other hand, in this paper, we propose the recognition of gesture actions as command-input-means. The user of client needs merely the camera (command-input-device), data-acquisition-device and a monitor. Data-acquisition-device consists of a scanner, the analog-digital converters, the minimum data memory and the "interface of communication via wire or wireless". Because it can be embedded to the portable device, the monitor should be type of a flat-panel-display.

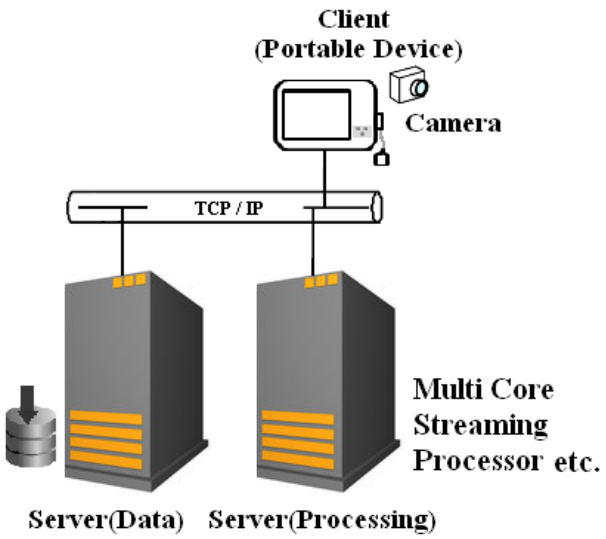


Figure 1: System Organization

2.2 Data Flow

The portable devices record data onto their storage memory and transmit the data to the server by "Transmission Data" command. The data flow of portable device is shown in Figure 2. The sensor (such as a camera sensor, a time of flight measurement sensor… etc) is set to the device internally or externally. The sensor (in the experiment, a camera is used as a sensor) takes the image information in order to control the device by gesture actions. Kinds of commands are I/O (such as data acquisition, monitoring), data transmission, etc. The processing of image recognition classifies gesture actions to each command type. These commands control sensors and actuators of the portable device. In other words, the gesture actions make the event for controlling the portable device. Especially, the event of Data-transmit transmits the acquired data by scanner from the client to Processing-Server. After the processing of Processing-Server is finished, the client receives the image data from Server via LAN or WLAN.

All of the data flow of system is described as follows.

1. If the operator grips the scanner, the client starts acquisition of the data and stores the data to Buff Memory in real-time.
2. If the command of Transmit Data is recognized by Image Recognition Unit, the data is transmitted from client to Processing-Server. After Processing-Server finished the reconstruction of a volumetric image from the data, the volumetric image is transmitted from Processing-Server to client via LAN or WLAN
3. Finally, when the command of Display Image is issued, display i/f outputs the volumetric image data to the monitor of the client.

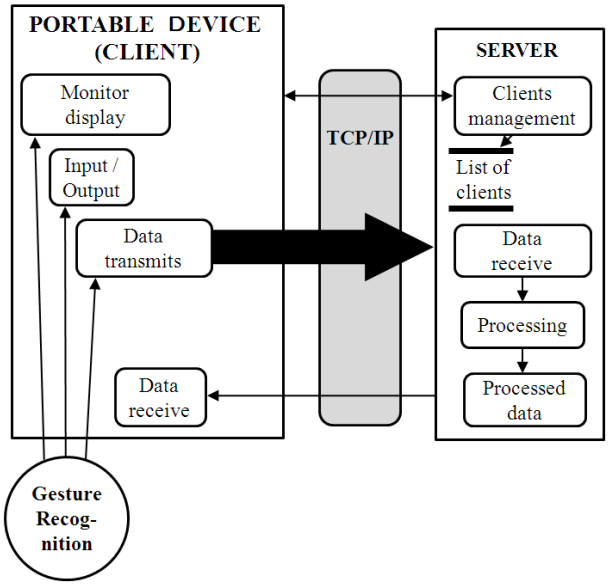


Figure 2: Data Flow of System

2.3 Commands Based on Gesture Actions

We show the commands prepared to clients in Table 1. As the mentioned in the section 2.2 of data flow, various kinds of command are prepared in order to control the portable device. One of commands is used for acquiring data. The operator grips a scanner in a hand for data acquisition. And other command is used for transmitting data. Commands for monitoring are used in order to display an image or information on the screen of monitor.

The operator must grip a scanner in a hand for data acquisition. If he communicate with machines via handheld devices like keyboard, mouse, laser-tracked pen and stylus on another hand, he cannot have the one any more.

Table 1: Assigned commands to gesture actions

Commands	Gesture Style	Direction
Acquire Data	Gripping Scanner	Client Only
Transmit Data	Assigning one finger	Client to Server
Display Image	Assigning two fingers	Server to Client
Display Information	Assigning three fingers	Server to Client

Therefore we use gesture recognition to operate the portable device. Object recognition in computer vision is the task of finding a given object in an image or video sequence. Humans recognize a multitude of objects in images with little effort, despite the fact that the image of the objects may vary somewhat in different viewpoints, in many different sizes, scale or even when they are translated or rotated. Objects can even be recognized when they are partially obstructed from view. This task is still a challenge for computer vision systems in general. However, we challenged to adapt the object recognition for the portable device.

2.4 Wireless Communication

The portable device as a client enables communication via wireless too. Wireless portable-device has many advantages such as power consumption, size, low cost, scalability, mobility and flexibility.

We show the organization of wireless system in Figure 3. The system contains servers and portable data acquisition device. Servers are laid to fixed location. Processing-Server responds to requests from the portable device and reconstructs an image from the acquired data by the portable device laid in the operation room. The server will return a reconstructed image to the portable device after the processing. The wireless communication provides the more flexibility to the distributed system.

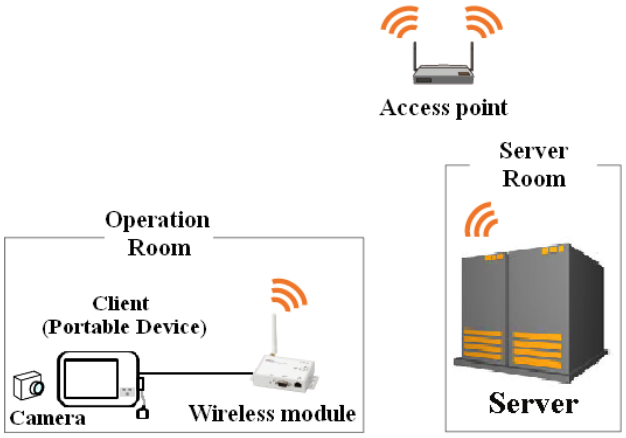


Figure 3: Wireless System Organization

3 DEVELOPMENT OF A PROPOSED PORTABLE DEVICE

3.1 Functional Prototype Device

We show the structure of client in Figure 4. This is functional prototype. The client consists of a "camera as an input-device", a data-acquisition-board, a monitor and a "scanner for scanning the object". The flat-panel-display-device is used for monitoring the image, but not yet embedded in portable device. The data-acquisition-board has communication ports via wire and wireless. Therefore, the client can work such as a standalone device without laying the nearby server. Furthermore, the operator of client can easily control the device by gesture actions.

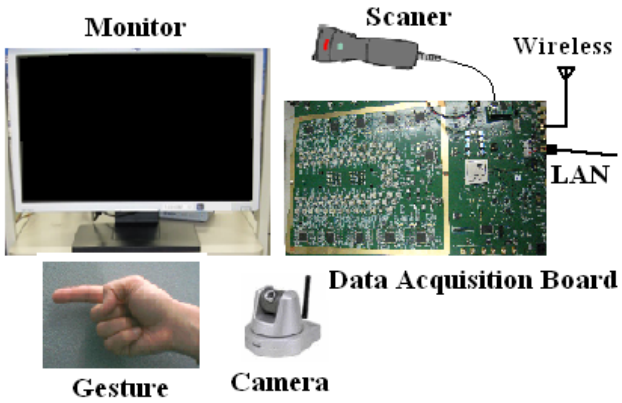


Figure 4: Functional prototype of the portable device

We show the functional block diagram of the data-acquisition-board in Figure 5. Data Acquisition Board has A/D Converters, Buff Memory (16Bits × 1K), CPU Core, Image Recognition Unit, LAN i/f and WLAN i/f. If the operator grips the scanner, the board starts the acquisition of the data and stores the data to Buff-memory in real-time. If the command of "Transmit Data" is recognized by Image Recognition Unit, the data is transmitted to Processing-Server. After Processing-Server finished the processing in order to reconstruct an image from the data, the board receives the image data from Processing-Server via LAN or WLAN. When the command of "Display Image" is issued, Display i/f outputs the image data to the monitor.

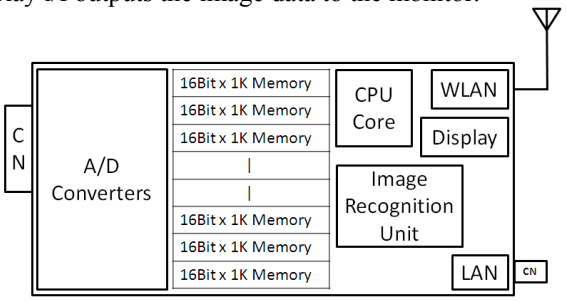


Figure 5: Functional block diagram of the board

3.2 Future Structure

We show the more suitable structure in Figure 6 for a client. The client will consist of a "camera as input-device", a monitor and a "scanner for scanning the object". The scanner will have the data acquisition function and the WLAN function. The WLAN inside of scanner will be used for communication with servers. While a camera and a monitor are separated in this figure, these devices will be put in the same case in future.

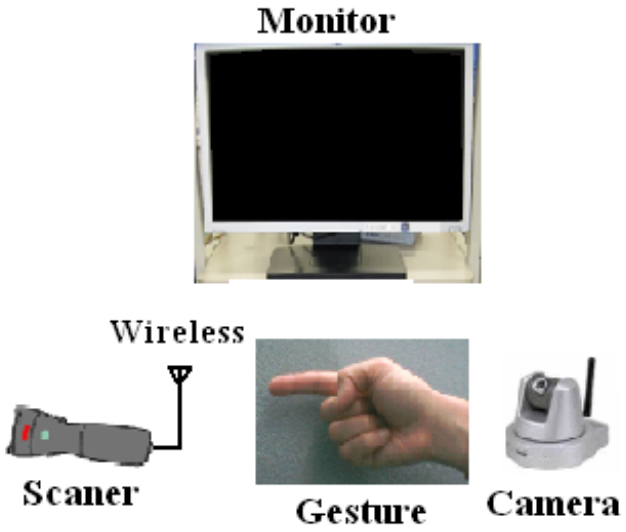


Figure 6: Functional parts for the suitable client

3.3 Gesture Recognition

We use histogram of oriented gradient descriptors in order to recognize gesture actions. Histogram of oriented gradient descriptors is feature descriptors used in computer vision and image processing for the purpose of object detection [10]. The technique counts occurrences of gradient orientation in localized portions of an image. This method is similar to that of edge orientation histograms, scale-invariant feature transform descriptors, and shape contexts, however differs in that it on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for improved accuracy. The essential thought behind the histogram of oriented gradient descriptors is that local object appearance and shape within an image can be described by the distribution of intensity gradients or edge directions. Since the histogram of oriented gradients descriptor operates on localized cells, the method upholds invariance to geometric and photometric transformations such changes would only appear in larger spatial regions. Moreover, "coarse spatial sampling, fine orientation sampling, and strong local photometric normalization" permits the "individual hand movements of operators" to classifying a position of finger roughly.

We recorded the 2D image of gesture action with a camera. Then, the image was saved to the memory. We created the template pattern from it. We input total images from a camera to classify images. We processed the recognition of pattern in the image by using histograms of oriented gradient descriptors. We show this result in Figure. 7. The rectangle of a black line is drawn around a recognized pattern.

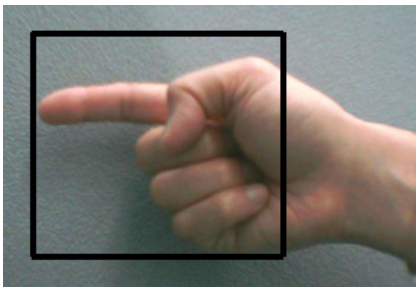


Figure 7: Example of Image Recognition for Classifying Gesture Actions

4 EVALUATION

4.1 Evaluation System

Client configuration is as follows. (Refer to Figure 4)

The client consists of a "camera as input-device", a data-acquisition-board, a monitor and a "scanner for scanning the object". The flat-panel-display-device is used for monitoring the image, but not yet embedded in portable device.

Server configuration is as follows.

Server has 4-core Xeon processor x 2 and GPU. We show the specification of a video card using GPU (nVIDIA Quadro FX4800) in the following.

- Chipset:Quadro FX4800
- Stream Processors: 192
- Memory Bandwidth: 76.8GByte/sec

The processor chip of video card (GPU) has the architecture of parallel streaming. The GPU has 192 streaming processors. If GPU is used in the server, the calculation speed will be able to become more than 100 times compared with a single-core processor. We used two cases of 19-inch rack-mount with 36 units. The size of server was extremely downsized by GPU. However, Servers are too big in the operation room yet.

4.2 Experiment and Results

We selected the processing of Optoacoustic imaging for the system evaluation. Optoacoustic imaging recently became subject of intense research in the field of medical diagnostics. We took the image by using the distributed system including the portable device. As light source in the near infrared and visible wavelength range are commonly used. The temporal shape and the duration of the ultrasound yield information about the distribution of absorbed optical energy. The ultrasound is measured outside the tissue and is used for generating tomography images of absorbing structures inside biological tissue [11]. The method of Optoacostic signal detection is shown in Figure 8 [12].

Optoacoustic imaging is a three-dimensional (3D) imaging technique since the scattered incident light pulse generates thermo elastic pressure waves in absorbing structures located inside the whole irradiated volume. The detected Optoacoustic pressure waves contain this 3D information which is utilized to reconstruct the distribution and shape of the absorbers in the irradiated volume. To acquire the amount of information necessary to reconstruct a 3D image, a method has to be used that efficiently samples the acoustic waves. One method is the measurement of two-dimensional (2D) snapshots of the pressure distribution in a plane at different delay times after the incident laser pulse. Based on such 2D measurements the pressure source can be reconstructed by using a radial back projection of the recorded pressure distributions from the detector plane into the source volume [13]. This is a very robust algorithm, which gives an image even if the number of collected 2D stress images and their resolution are low. Optoacoustic imaging has to solve optoacoustics wave equations. We need huge power to solve these equations. These equations are defined by equation (1), (2) and (3) [14].

Wave equation for the velocity potential ψ can be expressed by next equation.

$$\Delta\psi - \frac{\partial^2\psi}{c^2\partial t^2} = \frac{\beta}{\rho C_p} S \qquad (1)$$

Where grad ψ is the velocity vector of the wave motion, c is the speed of sound, β is the thermal expansion coefficient, ρ is the density, C_p is the specific heat capacity at constant pressure and S is the heat generated by absorption of an electromagnetic wave per unit volume and time. This equation is valid under the assumptions that heat conduction and viscous damping can be neglected. Under short pulsed irradiation the heat source term S can be expressed as $S(r,t) = W(r)\delta(t)$, using the volumetric energy density W.

With $p = -\rho\frac{\partial\psi}{\partial t}$, the wave equation for the acoustic pressure p can be expressed by next equation.

$$p(r,t) = \frac{\partial}{\partial t}\left[\frac{\iint p_{0(r')}ds}{4\pi c^2 t}_{|r-r'|=ct}\right] \qquad (2)$$

$$\text{With } p_0(r') = \frac{\beta c^2}{C_p} W(r')$$

The integration is performed over the surface of a sphere with surface element ds and radius ct around the detection point r.

In the following we will use a Cartesian coordinate system and assume that a line detector of infinite length is oriented parallel to the z-axis. It receives signals given by next equation (3).

$$q(x,y,t) = \int_{-\infty}^{+\infty} p(x,y,z,t)dz \qquad (3)$$

We evaluated the reconstructing of a volumetric image. We reconstructed image of a black colored sphere shaped object in an agar gel. We show the arrangement of target object in Figure 9. We performed a processing to reconstruct $64 \times 64 \times 128$ volumetric pixel imaging. We show the reconstructed 3D image in Figure 10 by using the reconstruction algorithm as mentioned above. In this case, it took about 60 mS to reconstruct a volumetric image by using 4-core Xeon processor x 2 and GPU in Server.

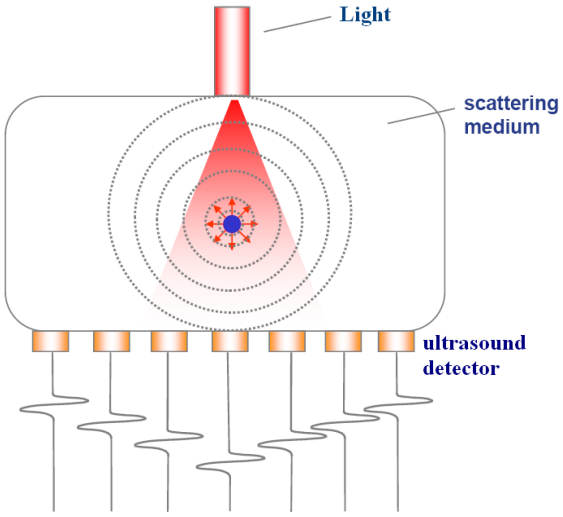


Figure 8: The method of Optoacostic signal detection

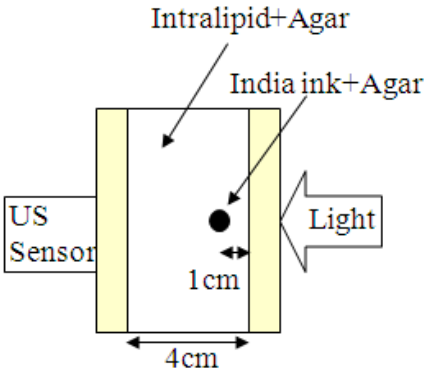


Figure 9: Arrangement of the target object

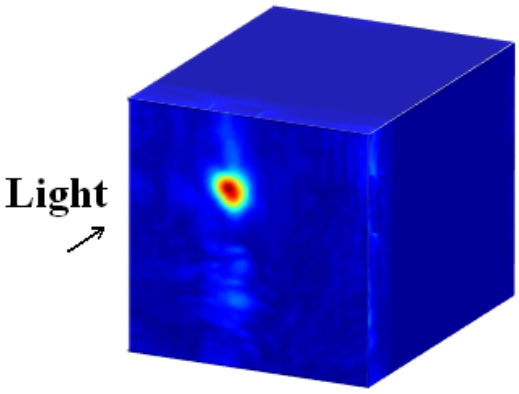


Figure 10: Reconstructed 3D Image

4.3 Summary of the Evaluation

Summary of the evaluation is as follows.

(1) The prototype of client is not small enough to treat it as a portable device. Our future target of the device-size is iPad-size including an embedded camera.

(2) The gesture recognition has sometimes missed to classify gesture actions. The rate of misrecognition is about 42%.

(3) It took 134 mS to process an Optoacoustic image by the system. The processing time is the total time of data acquisition time, image reconstruction time and communication time. It took about 60 mS to reconstruct a volumetric image by using GPU in Processing-Server. Processing speed by Processing-Server is preferable. However, it is not enough to reconstruct in real-time (about 16.7ms), we are trying to speed up the reconstruction time. The portable device has a 1Gbps Ethernet port. However the average speed of communication is slower than the speed of specification. It took about 74 mS to communicate between the client and the server. The transfer rate was about 122Mbps.

In case of wireless communication, the portable device has an 802.11n wireless port. However the average speed of communication is slower than the speed of specification. It took about 143 mS to communicate between the client and the server. The transfer rate was about 65Mbps.

We summarize these processing time in table 2.

Table 2: The time of Optoacoustic image processing

	Wired	Wireless
Amount time	134 mS	203 mS
Data acquisition time	100 μS	100 μS
Reconstruction image time	60 mS	60 mS
Communication time	74 mS	143 mS

5 CONCLUSION

In this paper, we proposed a distributed system with portable devices. The proposed system was Client-Server organization. Therefore, servers didn't disturb persons in the operation room, and persons around the client didn't feel the oppression. We developed a prototype of proposed portable device based on gesture recognition. We conducted the experiment by using a prototype of portable device. The device was operated by gesture actions. A hand of the operator could freely be used for other action. Furthermore, we conducted the experiment by using GPU. The server was downsized by GPU. And its calculation speed was faster than the conventional technology such as the CPU.

In the future tasks, we will design a smaller and thinner portable device. The communication and image-reconstruction speeds of the system will become more faster than results of this experiment. And, we will include more many cameras for the more accurate control of a portable device.

REFERENCES

[1] A. S. Tanenbaum, M. V. Steen, Distributed Systems: Principles and Paradigms, Pearson Education, (2003), T. Mizuno, T. Higashino, Y. Miyanishi, K. Suzuki, S. Nishiyama, F. Satou interpreted to Japanese.

[2] J. Kela, P. Korpipaa, J. Mantyjarvi, S. Kallio, G. Savino, L. Jozzo, S. Dimarca, "Accelerometer-based gesture control for a design environment," in Personal and Ubiquitous Computing, vol. 10, Springer-Verlag, London, 2005.

[3] R. M. Satava, "Future Technologies for Medical Applications," chapter in M. Akay, A. E Marsh, ds, blfornzatiori Technologies in Medicine, Volume I: Medical Siniulation and Education, New York: John Wiley and Sons, pp. 2 19-236, 2001.

[4] SARCOS, http://www.sarcos.com

[5] S. Barro, et al., "Intelligent Telemonitoring of Critical Care Patients", IEEE Engineering in Medicine and Biology, 18(4):80-88, 1999.

[6] B. Priddy, E. Jovanov, "Wireless distributed data acquisition system," System Theory, 2002. Proceedings of the Thirty-Fourth Southeastern Symposium, pp. 463 - 466 , 2002.

[7] A. R. Al-Ali, Y. R. Aji, H. F. Othman, F. T. Fakhreddin, "Wireless smart sensors networks overview," Wireless and Optical Communications Networks 2005. WOCN 2005. Second IFIP International Conference, pp. 536-540, 2005.

[8] C. Lethanhman, H.Isokawa, Kato, T, "Multipath Data Transmission for Wireless Thin Clients," Mobile Ubiquitous Computing, Systems, Services and Technologies, 2009. UBICOMM '09. Third International Conference, pp. 108 - 112 , 2009.

[9] J. D. Owens, M. Houston, D. Luebke, S. Green, J. E. Stone, and J. C. Phillips, "GPU Computing," Proceedings of the IEEE, Vol. 96, Issue 5, pp. 879 – 899, 2008.

[10] N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005. Vol. 1, pp. 886 – 893, 2005

[11] K. P. Köstli, D. Frauchiger, J. J. Niederhauser, G. Paltauf, H. P. Weber, and M. Frenz, "Optoacoustic Imaging Using a Three-Dimensional Reconstruction Algorithm," IEEE Journal of Algorithm Selected Topics in Quantum Electronics, Vol. 7, Issue 6, pp. 918 – 923, 2001.

[12] K. P. Köstli, M. Frenz, H. P.Weber, G. Paltauf, and H. Schmidt-Kloiber, "Optoacoustic tomography: Time-gated measurement of pressure distributions and image reconstruction," Appl. Opt., vol. 40, pp. 3800–3809, 2001.

[13] G. Paltauf, R. Nuster, M. Haltmeier and P. Burgholzer, "Experimental evaluation of reconstruction algorithms for limited view photoacoustic tomography with line detectors," April 2007.

[14] Ton G. van Leeuwen, Wiendelt Steenbergen, Srirang Manohar Johan van Hespen, "Biomedical Optics: Exploiting Light in Living Tissue,"

Session 2: Groupware & Education Systems
(Chairs Tomoo Inoue and Takuya Yoshihiro)

Research on Collaborative Learning Technology
－Recommended Study Materials to Overcome the Weak Point－

Yuji Wada [*], Yuma Hamadume, Shinichi Dohi [*], and Jun Sawamoto [**]

[*]Graduate School of Information Environment, Tokyo Denki University, Japan
[**] Faculty of Software and Information Science, Iwate Prefectural University, Japan
[*] yujiwada@sie.dendai.ac.jp, [**] sawamoto@iwate-pu.ac.jp

Abstract - Recently, e-learning has been spreading in educational institutions and corporate training. This technology tends to reduce the educator's load by the use of tools such as LMS (abbreviated e-Learning Management System). On the contrary, this technology does not necessarily reduce the load for the learner. For example, the repetition of online quizzes and the inspection of a great number of web teaching materials can become a load on the learner. Therefore, the collaborative learning recommendation system that recommends teaching material for overcoming the weak areas of specific learners was developed in the present study. This paper describes the system and the outcome of an experiment to evaluate this system.

Keywords: e-learning, collaborative learning system, recommend technology, weak point

1 INTRODUCTION

In many institutions of higher education and in corporate training, e-learning has been actively assimilated and developed. With a system [1] that partially uses an Adaptive Hypermedia System (AHS) to recommend a course of study suited to the learner, experimental results confirmed an improved learning effect, but the learning was based on online quizzes, which placed a burden on the learner.

Moreover, with a **bidirectional recommendation system** [2] that recommends highly relevant teaching material based on the learning history, it was confirmed that such highly relevant teaching material was recommended regardless of the ordering in which the teaching material was presented, and a questionnaire survey confirmed that the study time could be shortened. These findings demonstrate that recommended information should be suited to and not impose a burden on the learner.

Therefore, in this study, assuming that a higher access count in learning historical data implies an area in which the teaching material has a greater degree of difficulty, we developed a **collaborative learning recommendation system**[3] that recommends teaching material that is "material necessary for overcoming subject weaknesses" for learners exhibiting similar learning tendencies.

In this paper, we describe the experimental results of the collaborative learning recommendation system and propose a **similar preference data mining method** based on attribute data for new learners for which there is no historical learning data.

2 AIRS

An Individual Reviewing System (AIRS) [2] is an individual review support system that was developed in our institution (see Fig. 1). AIRS was developed to encourage review by generating instructional material (hereafter referred to as teaching material) from the learner's perspective and distributing easy-to-understand teaching material. The teaching material provided is generated based on lectures entitled "Basics of Information Processing" and "Database Systems" that are offered as courses at our institution. The teaching material is distributed in page units and displayed as web pages.

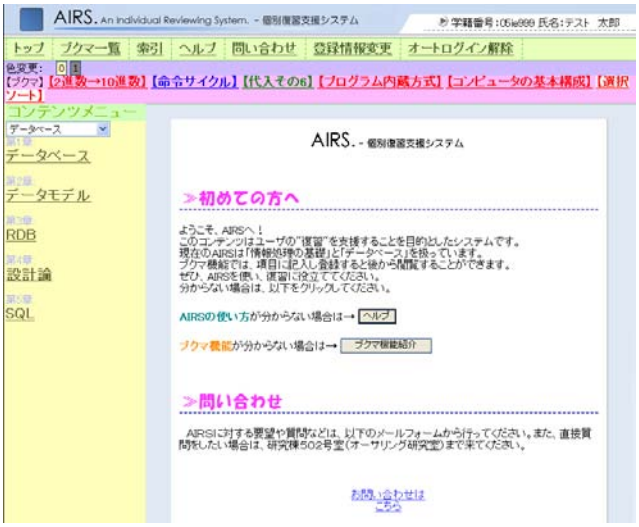


Figure 1: AIRS top page.

2.1 Teaching Materials

The teaching material of "What is a database?" in the lecture for database systems is shown in Fig. 2. As can be seen, explanations of keywords to be studied, the procedure for issuing a MySQL query, and other related information are displayed. A characteristic feature is that three types of methods of expression are provided for each set of teaching materials. Since different learners find different expressions to be easy-to-understand, various expressions are generated

by utilizing concise expressions, detailed expressions and expression through movement.

[Database]

What is a database?

A collection of the necessary data = data cluster

or

A data box into which data is put

A collection of data put together based on certain

Learner database

No.	Name	School register	Grade	Activity group
1	T	IE	B	Baseball
2	S	ID	S	Swimming
3	K	IE	A	Basketball

Current level: 1

Let's add more detail!

Figure2: Teaching materials.

2.2 Learning History

This study uses the learning history accumulated by the AIRS e-learning system. As shown in Fig. 3, the learning history is comprised of an identification number (vector_id), the date and time (access time), the learner ID (user_id), and a teaching material identification number (content_id). The first line of the history indicates that on October 13, 2005 at 13:42:53, learner number ID0 accessed teaching material 43. Over the course of usage from 2005 through 2008, the learning history increased and now contains over 27,000 entries.

3 COLLABORATIVE LEARNING RECOMMENDATION SYSTEM

A characteristic feature of the proposed method is that it identifies learners exhibiting similar areas of weakness and recommends information for overcoming these weaknesses. Fig. 4 shows the overall flow. This method functions by recommending teaching material via a collaborative learning recommendation system in which the learning history is used as the basic data. For new learners having no learning history, recommendations are made using the attribute information.

vector_id	access_time	user_id	content_id
1	20051013134253	0	43
2	20051013134254	0	19
3	20051013134258	0	46
4	20051013134300	0	1
5	20051013134300	0	1
6	20051013134307	0	1
7	20051013134309	0	4
8	20051013134310	0	7
9	20051013134314	0	88

Figure3: Excerpt of the learning history.

3.1 Learner Access Count Data

First, as a pre-process, the learning history is converted to generate data indicating the number of times each learner accessed the respective teaching material. Fig. 5 shows an excerpt of this data. The first line indicates the number of accesses by learner number 0 and shows that content1 was accessed 53 times, content4 was accessed 49 times and content7 was accessed 40 times. In addition to providing directly observable values, this data is thought to reflect such characteristics as the learner's interests and level of proficiency. In the case of a system such as AIRS that aims to support learning, the learners' objectives are to prepare for and review lectures, but because the use of such systems thus far has been limited to times prior to examinations, the learner's objective can be considered to be review. In other words, this data is thought to indicate areas in which the level of understanding is inadequate and weaknesses exist.

3.2 Data Mining of Similar Learners

The access count data described in Section 3.1 is compared for all learners to mine data for learners having similar weaknesses, or in other words, similar learners. Fig. 6 illustrates the concept of similar learner data mining. Here, Learner A, as well as Learner B and Learner C, has a learning history. At this time, Learner A accessed the teaching materials of "Types of variables," "Basics of the 'while' statement" and "Basics of the 'if' statement," Learner B similarly accessed the same three materials and Learner C accessed the materials of "Types of variables," "Basics of assignment" and "Functional dependence." In this case, because the teaching materials accessed by Learner A and Learner B are similar, there is a high degree of resemblance between Learner A and Learner B, and therefore they are similar learners.

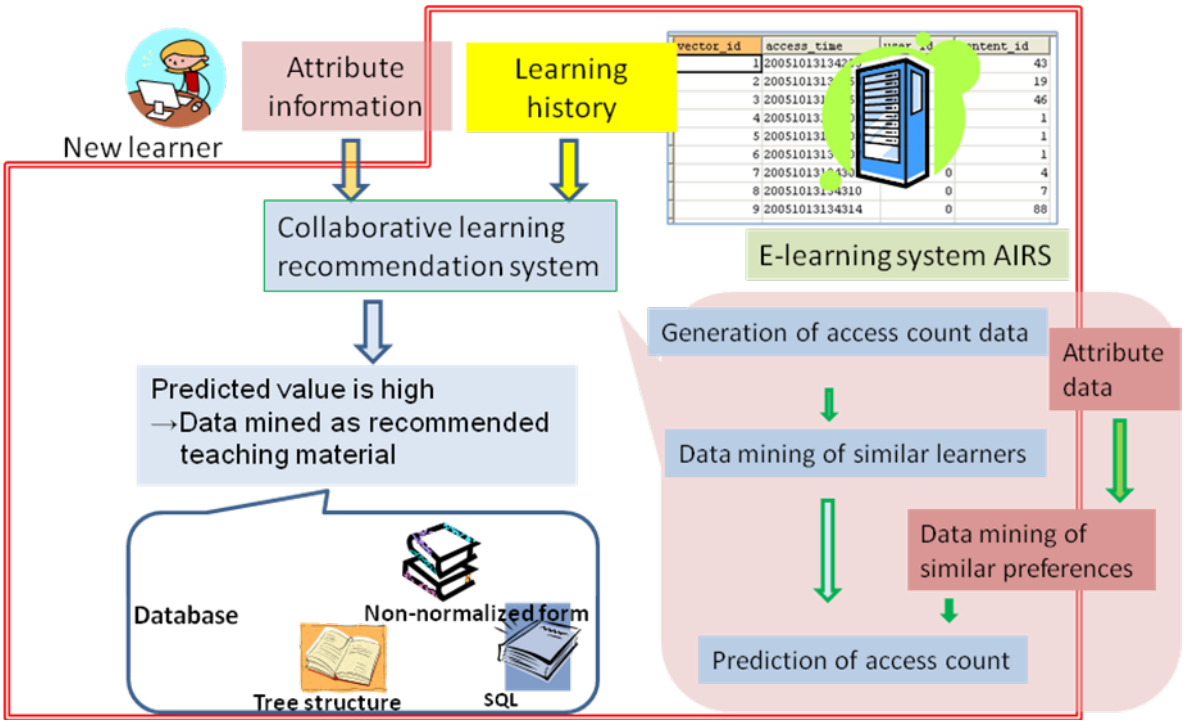


Figure 4: Overall system flow.

user_id	con1	con2	con3	con4	con5	con6	con7
0	53	0	0	49	0	0	40
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	2	0	0	1	0	0	2
4	9	0	0	7	0	0	5
5	5	0	11	1	0	0	0
6	4	0	0	0	0	0	1
7	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0
9	9	0	0	12	0	0	1
10	0	0	0	0	0	0	0

Figure 5: Access count data.

A correlation coefficient algorithm [4], implemented, for example, in collaborative filtering, is used in the actual mining of data for similar learners. The form of the computational equation applied in this research is shown in equation (1) below.

$$R_{ab} = \frac{\sum_{i=1}^{I} (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^{T} (a_i - \bar{a})^2} \sqrt{\sum_{i=1}^{T} (b_i - \bar{b})^2}} \quad (1)$$

$\mathbf{R_{ab}}$ is the resemblance between Learner A and Learner B. R is the first letter of the word resemblance. Values of R range from 1.0 to –1.0, and values approaching 1.0 indicate greater resemblance, values approaching –1.0 indicate less resemblance, and the value 0 indicates that there is no relationship. The numerator indicates covariance, and the denominator indicates the product of the standard deviations.

a_i indicates the number of times Learner A has accessed the i^{th} teaching material, and $\bar{a}$ indicates the average number of accesses per teaching material. T indicates the total number of teaching materials. The learners found to have a high degree of resemblance according to this method are mined as similar learners.

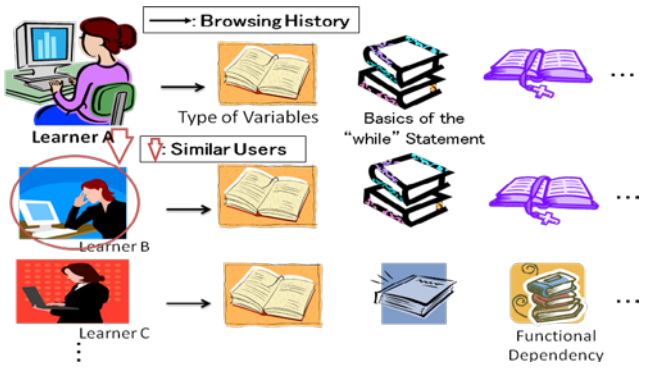


Figure 6: Method of mining data for two similar learners

3.3 Data Mining of Recommended Teaching Material

Here, assuming that the history of similar learners contains a history of overcoming weaknesses, teaching material that has been accessed many times by similar learners is mined as teaching material that helps to overcome these learners' weaknesses. Fig. 7 shows the method of mining recommended teaching material. In this example, similar learners accessed "Types of variables" three times, "Basics of the 'while' statement" nine times, "Basics of the 'if' statement" five times and "Basics of arrays" two times. If recommended teaching materials are

mined from the historical data, "Basics of the 'while' statement" and "Basics of the 'if' statement," which were the most frequently accessed, will be recommended to learners.

The actual computation for mining recommended teaching materials is performed by a method [4] that calculates the predicted value used with collaborative filtering. The data for this calculation is based on the resemblance among similar learners and the learner vector of similar learners, and the value calculated is the predicted value of the access count. In this study, the predicted value is computed using equation (2) below.

$$P_{a,1} = \bar{\alpha} + \frac{\sum_{U_i \in User} (C_{i,1} - \bar{C}_i) r_{ai}}{\sum_{U_i \in User} |r_{ai}|} \quad (2)$$

Here, $P_{a,1}$ is the predicted value (predicted access count) necessary for Learner A to overcome a weakness in teaching material 1. For example, in the case where $P_{a,1}$ has a value of 3.4, Learner A is thought to be able to overcome his or her weakness in teaching material 1 by accessing that content approximately three times. Also, $\bar{\alpha}$ is the average number of times that a group of learners having some correlation with the earner A (i.e., learners having a non-zero resemblance value) accessed the teaching material; in other words, the average of the average access count. "User" is a group of learners having some correlation with Learner A, and the i[th] learner within this group is denoted as U_i. $C_{i,1}$ indicates the number of times that the i[th] learner accessed teaching material 1, $\bar{C}_i$ is the average access count and indicates the average number of times that the i[th] learner viewed any single teaching material, and r_{ai} indicates the resemblance between Learner A and Learner B. In this manner, teaching material for which the predicted access count is a large value is mined as recommended teaching material.

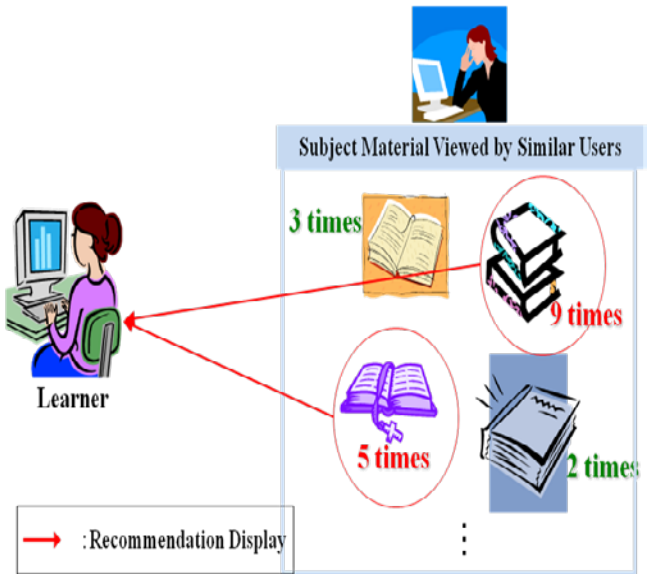


Figure 7: Method of mining data for recommended teaching material.

4 EXAMPLE OF COLLABORATIVE LEARNING RECOMMENDATION

4.1 Analysis of Learning History

First, a graph of the distribution of 368 learners according to the total access count is shown in Fig. 8 below. Recommended teaching materials were mined and analyzed for these learners.

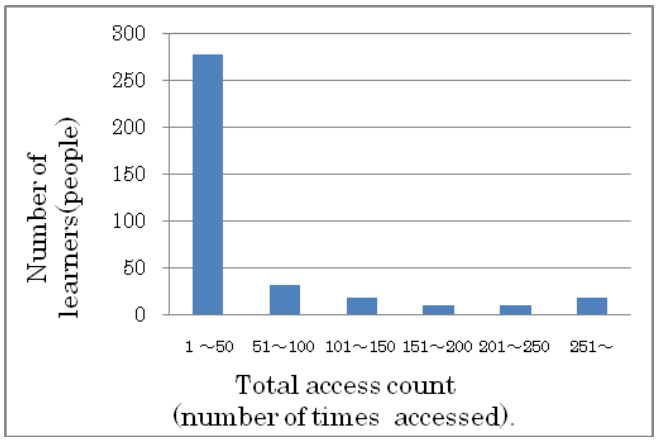


Figure 8: Distribution of learners.

4.2 Comparison of Recommended Teaching Materials

For the purpose of comparing recommended teaching materials for each learner, four learners were selected from the learning history of the 368 learners characterized in Fig. 8. The results show that two learners made between 1 and 50 accesses (user_id=10,200), one learner made between 101 and 150 accesses (user_id=336) and one learner made between 201 and 250 accesses (user_id=427). Table 1 lists the recommended teaching materials for these four learners. Here, recommended teaching materials 1 to 4 are arranged in their perceived order of effectiveness in overcoming weaknesses, and the numbers enclosed in parentheses indicate the number of times of prior access. Here, focusing on recommended teaching materials 1, 3 and 4, it can be seen that no recommended teaching materials are common among the learners. Accordingly, there is believed to be a high probability that teaching material suited to the learners will be recommended (effective level (1)).

Also, as can be seen from the three users other than user_id427, the numbers enclosed in parentheses are all 0. In other words, the recommended teaching materials have never been accessed before. As a result, teaching materials that had previously been unknown are recommended to the learners (effective level (2)).

4.3 Overall comparison of Recommended Teaching Materials

For effective levels (1) and (2) of Section 4.2, we compared and verified the number of occurrences of recommended teaching material 1 which, based on data

from all learners, had the highest predicted values. The results are shown in the graph of Fig. 9.

In evaluating the likelihood that the recommended teaching material will be suited to the learners (effective level (1)), the bias in the recommendation toward some of the teaching materials was found, and this is a problem. The 136 instances of teaching material 3 "Databases" were found. In other words, the same teaching materials are recommended to approximately 35% of the 368 learners. An analysis of the data of the 136 learners revealed that 130 of them had a total access count of 10 or less. Consequently, learners having low access counts are considered to have unclear learning characteristics, and therefore "Databases," which has a statistically high access count, is recommended for these learners.

Also, "Variables" had a total occurrence rate of 9% (33 incidences) and "Stored program concept" had a total occurrence rate of 16% (59 incidences). Bias in the occurrence rate of these teaching materials was not due to the access counts. Among other recommended teaching materials, those having the highest occurrence rates occurred approximately 20 times, and there was an overall 5% overlap, but this is thought to be a relatively small value. Furthermore, of the 238 learners not having a low access count history, teaching materials other than the three biased recommendations discussed here were recommended for 61% (146 learners). This finding is considered to show that, at present, the recommendations are generally suited to the learners.

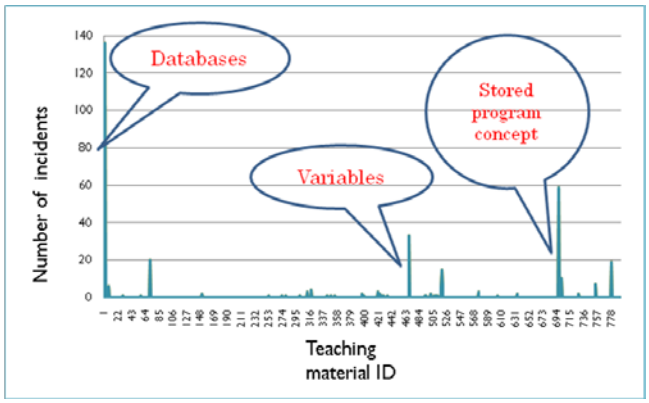


Figure 9: Statistical comparison of recommended teaching material 1.

However, regarding the point that unknown teaching material is being recommended (effective level (2)), historical access counts of recommended teaching materials were compared using the data for all learners. The results showed that 246 of the 368 learners (approximately 67%) had a history of 0 accesses for all recommended teaching materials 1 to 4. Also, among the total 1476 (= 368×4) recommended teaching materials, 1302 (approximately 88%) had a historical access count of 0 times. This finding is considered to show that many learners receive recommendations to use unknown teaching materials.

Table 1: Comparison of recommended teaching material for four learners

user_id	10	200	336	427
Recommended teaching material 1	Linear search (0)	Hash collision (0)	Variable (0)	Array (0)
Recommended teaching material 2	Stored program concept (0)	Chain method (0)	Complexity of bubble sorting (0)	Stored program concept (0)
Recommended teaching material 3	Control unit and database unit (0)	Open address (0)	Characteristics of stored program concept (0)	Database (5)
Recommended teaching material 4	Byte (0)	Logic data model (0)	Complexity of insertion sort (0)	3-level schema (7)

5 SIMILAR PREFERENCE DATA MINING METHOD

For new learners who do not possess a learning history, a method is proposed for using attribute data to mine data for learners having similar preferences.

5.1 Attribute Data

Here, learning preferences and information, such as the semester indicating the learning time period, are acquired from personal information of age and gender, hobby preferences indicating an individual's interests, and learning-related information such as the learner's proficient subject, weak subject, average study time.

5.2 Mining of Data for Learners Having Similar Preferences

Fig. 10 shows the concept of mining the attribute data for learners having similar preferences. Here, the new learner's weak subject is information, the proficient subject is English, and their hobby is swimming; Learner D's weak subject is information, proficient subject is English, and hobby is baseball; the learner E's weak subject is English, proficient

subject is information, and hobby is astronomical observation. At this time, new learners and Learner D are considered to have similar preferences since their learning preferences and hobby preferences are similar.

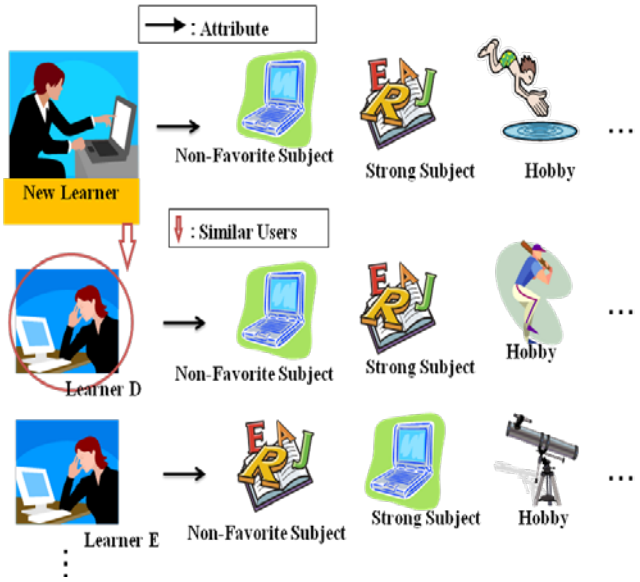


Figure10: Method of mining data for learners having similar preferences.

5.3 Organization of Attribute Data

Organizing the attribute data allows interrelations among the attribute data to be expressed. To display a large amount of attribute information and to clarify the learner characteristics, widely known dictionaries were consulted [5]. Fig. 11 shows a portion of a graph that organizes attributes related to hobby preferences. In the figure, the root nodes are hobby preferences, sub-nodes are genres, and at deeper levels, the genres are subdivided further. For example, "Sports" on the 1st level is widely subdivided on the 2nd level into activities such as swimming, fishing, track, skiing and so on, and swimming is further subdivided into synchronized swimming, swimming racing, water polo, etc. Then, to complete the attribute organization, swimming racing is finally subdivided according to the swim stroke (name of specific technique). With this hierarchy, attributes become more specific at deeper levels, and common attributes are thought to indicate a greater degree of resemblance. Here, level 1 is assigned 0.5 points, level 2 is assigned 1 point, level 3 is assigned 2 points and level 4 is assigned 4 points. Also, water polo is listed as both a ball game and swimming. In the case of attributes such as these, regardless of the event, relationships are clarified so as to facilitate the selection process.

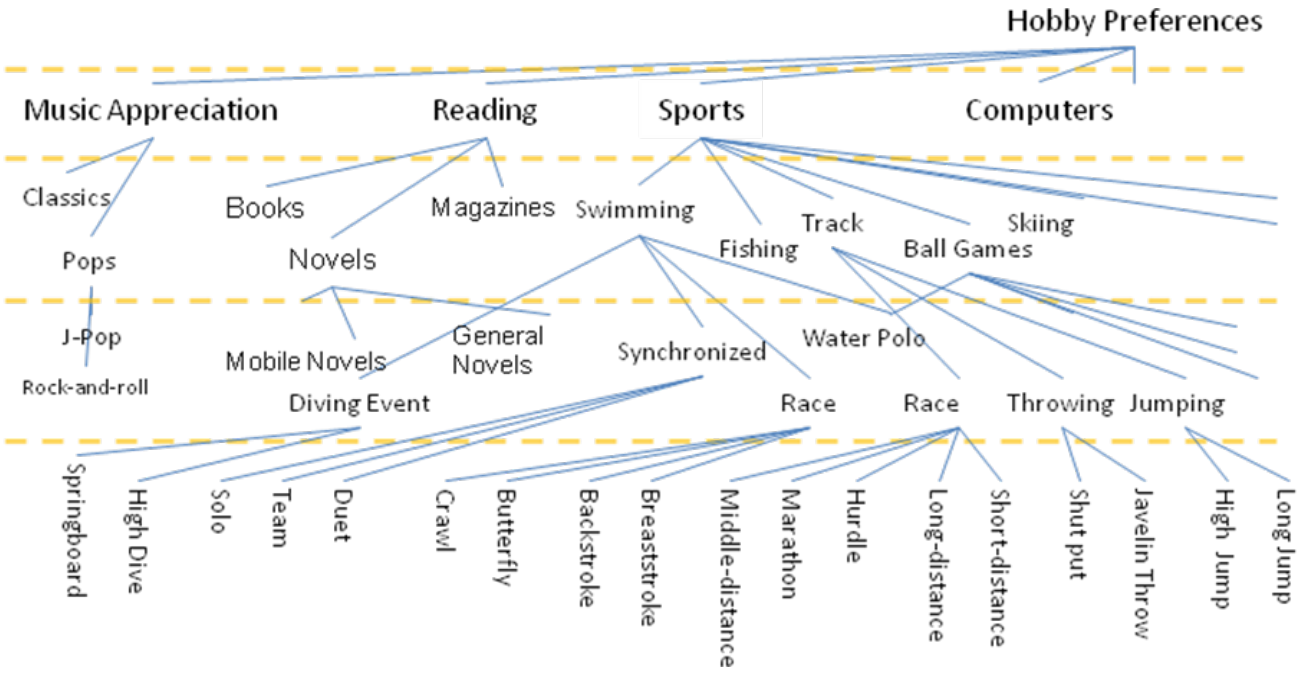


Figure 11: Organization of hobby preference data.

5.4 Acquisition of Hobby Preference Data

The learners select attributes from among the organized items and this information is stored in a database and used to generate a table in which the attribute data expresses learner characteristics. Fig. 12 shows refinement of most attribute data via the attribute data, input interface, a table expressing attribute data characteristics, and the method for calculating the degree of resemblance. For example, in the case where hobby preference → Sports → Swimming → Breaststroke is selected from the input interface, data for the table is generated as Fishing 0, Swimming 1, Crawl 0, Breaststroke 4, Skiing 0 and Sports 0.5.

5.5 Calculation of Similar Preferences

Using the table obtained in Section 5.4, the method for identifying new learners with similar preferences is described below. First, in the case of a new learner and Learner A, due to their similar interest in sports, a degree of resemblance of 0.5 points is assigned. Next, the new learner's and Learner B's similar interest in the breaststroke is assigned a degree of resemblance of 4 points, and the new learner's and Learner C's similar interest in swimming is assigned a degree of resemblance of 1 point. In this manner, similarities appearing at deeper levels are assigned higher degrees of resemblance. Also, fishing and swimming, which at first glance do not appear to be related, can also be said to have a degree of resemblance since they share a similar genre of sports. Recommendations are made to new learners based on the degree of resemblance calculated in this manner and the recommended teaching material mined in Section 3.3.

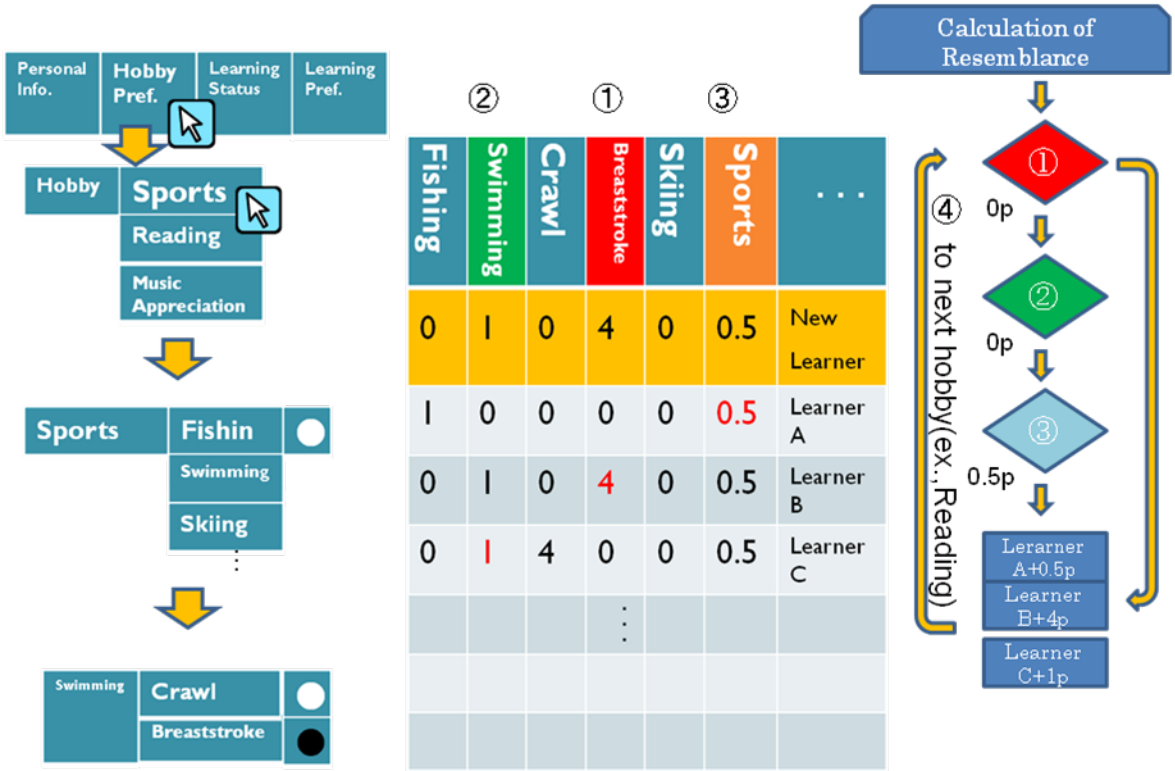


Figure 12: Refinement of attribute data via the attribute data input interface, the table expressing attribute data characteristics, and the method of calculating the degree of resemblance.

6 CONCLUSION

In this study, a system was developed for recommending teaching materials that help learners overcome their weaknesses. In experiments, we mined the teaching materials recommended material for each learner, and indicated the likelihood with which recommendations suited for the learner and recommendations of unknown teaching material would be made. However, proof of whether the teaching material helps to overcome areas of weakness, which is the objective of this study, could not be derived from the prior learning history.

In the future, proof that the teaching material contributes to overcoming weaknesses will actually require the use of a recommendation function, followed by a questionnaire survey and an analysis of the learning history. Also, regarding the method for making recommendations to new learners, the organization of attribute information and an algorithm are proposed. Then, we plan to implement a recommendation algorithm for new learners.

ACKNOWLEDGMENT

We would like to express our deepest gratitude to the teachers, researchers, and students who kindly participated in the experiments on our collaborative learning recommendation system with AIRS.

This research is supported by a Grant-in-Aid for Scientific Research C (Subject No. 21500908: Research on Adaptive Recommendation Technology based on Bidirectional Recommendation Technology for E-Learning Texts).

REFERENCES

[1] E. Triantafillou, A.Pomportsis, S. Demetriadis and E.Georgiadou, The value of adaptivity based on cognitive style: an empirical study, British Journal of educational technology, Vol.35, pp.95-106(2004).

[2] Y. Wada, S. Matsuzawa, M. Yamaguchi, and S. Dohi, Bidirectional Recommendation Technology for Web

Digital Texts, Journal of Digital Information Management, Vol.8, No.4, pp.240-246 (2010).
[3] Y. Wada, Y. Hamadume, S. Dohi and J. Sawamoto, Technology for Recommending Optimum Learning Texts Based on Data Mining of Learning Historical Data, International Workshop on Information (IWIN 2009), pp.143-150(2009).

Pictograph Communication using Tabletop Interface

Jun Munemori*, Takuya Minamoto*, Junko Itou*, Ryuki Sakamoto*

*Faculty of Systems Engineering, Wakayama University, Japan
{munemori, s105052, itou, rkskmt}@sys.wakayama-u.ac.jp

Abstract - The interpretation of the pictograph is approximately common through the world. We have developed a meeting type pictograph chat system that we assume to use at the tourist information center. The system can easily add a photograph as a pictograph, and use a tabletop interface for easy operation. Two users can operate the system simultaneously. Results of experiments show that the understanding degree of chat contents tended to improve when we used the photograph as a pictograph.

Keywords: pictograph, chat, tabletop interface, tourist support

1 INTRODUCTION

The Japanese who traveled in the foreign countries or the foreigners who traveled in Japan increased, and opportunities to regard communication as a foreigner increased. However, it is not still easy to communicate with a partner, who speaks a different language. Therefore a system supporting communication between different languages is requested without learning a specific language. The interpretation of the pictograph is approximately common through the world. Some systems support for communication using pictographs [1],[2].

We have developed a meeting type pictograph chat system that we assume to use at the tourist information center. The system can easily add a photograph as a pictograph, and use a tabletop interface (Diamond Touch Table) [3] for easy operation. In this paper, we show the experimental results using the system.

Chapter 2 explains the related work about this system, and Chapter 3 shows the proposed system. Chapter 4 describes the experiments and results, and Chapter 5 is the conclusion.

2 RELATED WORK

Pictograph chat communicator II [2] is a system, using only pictographs to communicate each other. The system consisted of two PCs via LAN. Pictographs are in color but a part is monochrome symbols. There are 547 pictographs in the system. This system has 9 tabs (including the History Tab), and pictographs are divided into 8 tabs. The application experiments of this system were performed by a Japanese student and a foreign student or an overseas international conference participant. The communication was performed smoothly. There was an opinion that there were few kinds of pictographs and few objective pictographs.

Zlango [4] is a pictograph-based system built for web and mobile messaging. The system has about 200 pictographs, which are changed from time to time, depending on its usage. Unused pictographs are deleted and new ones are being added to the system. The pictographs are divided into groups such as “People”, “ Actions”, “Places”, and “Feelings”. Zlango was developed in Israel and could be installed in cell phones in 12 countries. Zlango’s customers include Portugal Telecom/TMN, Globe (Philippines), Kiyv Star (Ukraine), Celcom (Malaysia) and other mobile operators.

3 THE PICTOGRAPH CHAT COMMUNICATOR III-T

3.1 Design policy

(1) The use of the tabletop interface

We use Diamond-Touch Table, which is one of a tabletop interface. The system has a large view area and enables the intuitive simultaneous operation with plural people.

(2) A function to display a proper noun as a pictograph

We implement a function, which adds photos or images as pictograph. The pictographs which are made from these photographs and images let difference in interpretation of the communication reduce.

3.2 System configuration

The system was developed for Diamond Touch Table. In the upper part of the system, a mirror reflects a picture from a projector, and then the picture is displayed on a display of Diamond Touch Table. The screen display domain is 490mm x 650mm. Two users use it in a meeting. They can operate the system simultaneously. They can use it synchronously. The number of pictographs is 560 and the size of pictograph is 45x45 pixel (29mm x 31mm on the screen). The system can use photographs as pictographs. A photograph can be fetched from Web or a digital camera. We use Flash, PHP, and XAMPP for the system. Figure 1 shows a screen of the system. The system consists of about 2,000 lines programs.

3.3 Operation of system

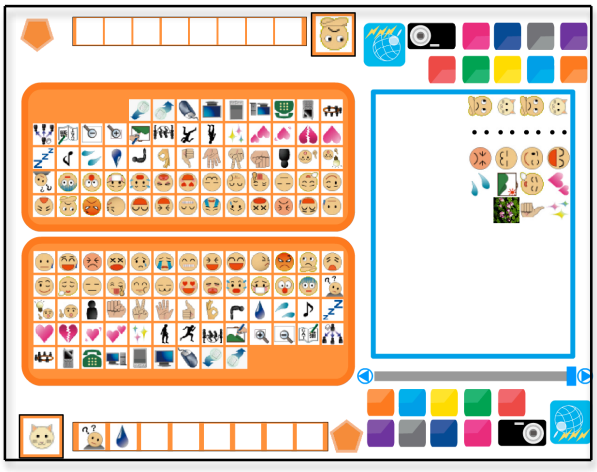


Figure 1: A screen of Pictograph chat communicator III-T.

At first the user chooses one's icon. In Figure 1, “cat” icon (left in bottom) is a user’s icon. A supporting pictograph (left in middle) is displayed when user chooses a tab (square and colorful button in bottom). If user finds the objective pictograph, then user touches the pictograph. The pictograph is selected and then input to the input field (left in bottom). In Figure 1, pictographs of “question” and “drop of water “ are in the input field. The function of delete and sorting are equipped at the input field. User repeats the operation mentioned above and makes a sentence. If the sentence is completed, then user touches the input button (pentagonal button in the middle of bottom). The sentence is represented vertically in the output field (right in top). Another user operates it in the same way in the other side simultaneously.

The additional function of the pictograph was implemented by PHP and Flash. When user wants to add the pictograph, user pushes the browser start button (a picture of the earth in the right of bottom) of Figure 1 and runs a browser. Next, user searches the image which user want to add as a pictograph on Web, and then user saves it in a designated folder manually. When user takes in a photograph from a digital camera and flash memory, user copies it in a designated folder and has only to save it.

4 EXPERIMENTS AND RESULTS

4.1 Experiments

The experiment went with two one set. One acts as a tourist, and the other acts as a member of explanation of the tourist information center at Osaka city.

We explained only the simple operation of the system to a subject. We do not explain the meaning of the pictographs. When user cannot express the content that user wanted to convey only with a pictograph prepared beforehand, we directed that user took an image from Web at any time. The experiment was finished when a member of explanation of the tourist information center replied it for all questions of the tourist.

The procedure of the experiment is shown below.

(1) The subject of the part of tourist asks questions decided beforehand such as a place or a route about sightseeing spot to a member of explanation position of the tourist information center. The questionary is shown Table 1. We do not tell a content of question to a member of explanation of the tourist information center at all. The subject assumed that there was knowledge around the sightseeing to some extent. Figure 2 shows a scene of experiment.

(2) The subject wrote down how he/she interpreted the meaning of one's remark and the remark of the partner after the chat end. The subject did questionnaire entry last.

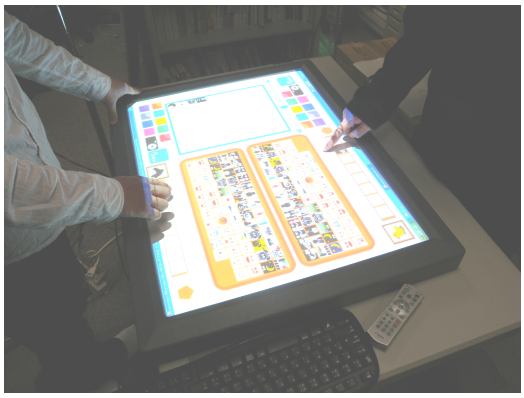


Figure 2: A scene of experiment.

Table 1: Questionary.

Questionary
（1） How do I go from Osaka to the Chinatown of Kobe?
（2） How does it take time?
（3） How much does the fare depend on?
（4） Which direction should I walk in if I arrive at the station?
（5） What is the recommended cooking in the Chinatown?

4.2 Results of Experiments

We have performed experiments using the system eight times. Three foreigners act as tourist (other five times were Japanese). Table 2 shows an experiment number and the profile (nationality, male/female, pictograph user/non-user) of the subject.

Figure 3 shows a result of a chat (experiment No.7). The bus icon corresponds to the tourist and the Penguin icon corresponds to the member of explanation of the tourist information center. The result that described the meaning of the remark by one's pictographs is shown below. The pictographs, which are made from pictures of a Chinatown, Chinese food, the station name board of Sannomiya Station, and the steamed meat bun, are used.

Table 2: The profile of the subject.

No.	Tourist	Pictograph	A member of a tourist information center	Pictograph
1	Japanese student —Male	User	Japanese student—Male	User
2	Japanese student —Male	Non-user	Japanese student—Male	User
3	Japanese student —Male	User	Japanese student—Male	User
4	Japanese student —Male	Non-user	Japanese student—Male	User
5	Foreign student — Female	User	Japanese student—Female	User
6	Japanese student —Male	Non-user	Japanese student—Female	Non-user
7	Foreign student — Female	User	Japanese student—Male	User
8	Foreign student — Female	User	Japanese student—Male	User

(1)
(2)
(3)
(4)
(5)
(6)
(7)
(8)
(9)
(10)
(11)
(12)
(13)
(14)
(15)
(16)
(17)

(1) Because I want to eat Chinese food in a Chinatown, how is it good in a train if I go?
(2) You can eat if you get on a train for Sannomiya Station.
(3) How does it take time by a train?
(4) It is 2:00 (misunderstanding).
(5) How much is a price to Sannomiya?
(6) It is 250 yen.
(7) Which direction should I go to from the station?
(8) You walk from Sannomiya.
(9) Which direction should I have walked in?
(10) That exists a direction, where many shops are existed, from the station.
(11) Then there is a Chinatown.
(12) Thank you.
(13) How does it take time on foot from the station to the Chinatown?
(14) You arrive on foot in five minutes.
(15) What kind of delicious dish is it eaten?
(16) A steamed meat bun is delicious.
(17) I like steamed meat buns.

Figure 3:Example of chat.

Table 3 shows time and output linage and the understanding degree that were needed for the chat of each pair. We take the evaluation method that Munemori [5] suggested in a calculation of the understanding degree into account.

Table 3: Time, output linage, and the understanding degree that we needed for a chat.

Mean value
Time (minutes): 23.3
Output linage: 19.5
The understanding degree of a tourist (%): 90.9
The understanding degree of a member of a tourist information center (%): 94.3
The understanding degree of all subjects (%): 92.9%

Table 4 shows the ratio of the image and the ratio of the image per line, and understanding degree of each experiment. The term of image includes photographs and images. “Image/line” means the number of images divided by the number of lines.

Table 4: Rate of image (%), image/line (%), and understanding degree (%).

No.	Rate of image (%)	Image /line (%)	Understanding degree (%)
1	9	29	89
2	6	22	88
3	14	50	91
4	15	67	93
5	2	5	95
6	11	32	93
7	15	69	96
8	19	77	98
Mean value	11.4	43.9	92.9

Table 5 shows the questionnaire results of experiments. The value of Table 5 shows the mean value of each 8 person about tourists or members of explanation of the tourist information center.

The description part of the questionnaire is as follows.

(1) About adding a photograph and an image to a chat

-Because it was fetched an image by Web, it is easy to have come to express it.

- Because a proper noun comes as well as a nuance properly, it is good.

- It is good that I can add it, but may be clogged up in a part of the letter input in the difference of the language when it is a foreigner handling Arabic with a Japanese PC.

(2) About system

- It was easy to take the pictograph big, but it was difficult for a lot of tabs to look for it.
- Not only I open up a browser, search results should be displayed in a chat screen.
- I want operation of the pictograph input to be possible with drug and/ or clicks.
- It is difficult to express progress of time.
- If there was a brief cartoon film pictograph, it was better.
-You should make a tab used well.
- There were many pictographs, which I did not use.
- I wanted the enlarged function of the photograph.
- I want relevance in the content of tab and the tab itself.
- It was difficult to show a direction.
- The destination of the image is hard to find.

Table 5: Questionnaire results of experiments.

Questions	A tourist	A member of a tourist information center
(1) Pictograph		
1. I was able to understand the meaning of all pictographs in the system.	3.8	3.6
2. Sentence making (using pictographs) was easy.	3.1	2.8
3. There were targeted pictographs (pictographs that I wanted to use).	3.9	2.7
4. I was able to understand the things my partner was trying to say.	3.6	3.5
5. I was able to make sentence what I want to say.	3.4	3.0
(2) Diamond Touch Table		
1. The input operation was convenient.	4.3	4.5
2. Elimination and sort of pictographs was easy.	4.3	4.1
3. Pictograph was easy to see.	4.0	4.1
4. I was able to look for the targeted pictographs smoothly.	3.3	2.0
(3) System		
1. Adding images to pictographs was convenient.	4.5	4.5
2. It was easy to understand the sentence when images were added to pictograph.	4.5	4.6
3. We can see the pictograph of our partner in input field. I predicted that the partner wanted to tell and was able to consider it to my sentence making.	3.6	4.0

１：I disagree strongly, ２：I disagree, ３：I don't agree nor disagree, ４：I agree , ５：I agree strongly

4.3 Discussion

(1) Pictograph chat

- About the interpretation of the pictograph

The understanding degree of chat was 93%. By the former experiments of general purpose, the understanding degree of chat was 91% and input line par minute was 2.3 [2].

-About Diamond Touch Table

The average of the evaluation to show in the following assumes it the mean of a tourist and the member of explanation of the tourist information center.

Question item (3) - 3 of Table 5 (We can see the pictograph of our partner in input field. I predicted that

the partner wanted to tell and was able to consider it to my sentence making.) was evaluated 3.8. During an experiment, a subject assisted an input of the partner, and a subject also replied it to see an expression of the partner. It may be said that there was a merit to make face-to-face communication by Diamond Touch Table.

(2) The consideration about the additional function of the pictograph

The average of the evaluation to show in the following assumes it the mean of a tourist and the member of explanation of the tourist information center.

Question item (3) - 1,2 of Table 5 (Adding images to pictographs was convenient and it was easy to understand the sentence when images were added to pictograph) was evaluated 4.5. As for the ratio of the image for pictograph and the ratio of the image for the output linage, understanding degree tended to become high if the ratio of the image was high. As for the coefficient of correlation, that there was equilateral correlation with 0.46, 0.51 was confirmed each. From the questionnaire description part above (verse 4.2), it evaluated good.

(3) Discussion about the experiment environment

About question item (1) – 3,5 of Table 5 (There were targeted pictographs that I wanted to use and I was able to understand the things my partner was trying to say), a tourist evaluated them 3.9, 3.4 each, and the member of explanation of the tourist information center evaluated them 2.7, 3.0. About question item (2) – 4 of Table 5 (I was able to look for the targeted pictographs smoothly), a tourist evaluated it 3.3, and a member of explanation of the tourist information center evaluated it 2.0. For both questions, the evaluation of the member of explanation of the tourist information center was low. Therefore, the expression of the answer side is more difficult than that of the question side. It is thought that pictographs specialized in a guidebook are necessary. There was no difference between the combination of the subject by this experiment.

(4) Discussion about the difference in interpretation

About the question item (4) of Table 1 (Which direction should I walk in if I arrive at the station?), a tourist asked as Figure 4, and the member of explanation of the tourist information center answered as Figure 5. Both subjects understood "which direction I should have gone" about the question of Figure 4. By the answer of Figure 5, there existed various answers like "go straight the hospital right", "go the hospital in the south", or "went along the bottom of the hospital". There existed some interpretation. We understood that we were hard to express an act to be twisted. It is difficult to express the act to be twisted in Figure 5. The pictograph, which is showed as "bending to the left" "bending to the right" is requested.

Figure 4: A question tourist.

Figure 5: The answer of the member of tourist information center.

5 CONCLUSION

We have developed a meeting type pictograph chat system that we assume to use at the tourist information center. Two users can operate the system simultaneously. We show the consideration of the result of the experiment below.

(1) It was easy to come to express a style by using a photograph and images such as a building and a place, and the scenery as a pictograph. Then, it was easy to have come to tell content to a partner. In addition, the average understanding degree was high with 93%, and understanding degree tended to become high if the ratio of the photograph and image was high.

(2) A pictograph specialized in a guidebook to "go to the north" and I "turn to the right" is necessary to do a guidebook.

(3) We were able to confirm that user predicted that the partner wanted to tell, because they met each other. So, user can support the input of the partner. This operation is one of the merits of the system.

In the future, improvement of the interface (ex. adding a map function), the making of a pictograph specialized in a guidebook, and an example sentence indication function will be requested.

REFEENCES

[1] Pangaean http://www.pangaean.org/

[2] M.B.Mohd Yatid, T.Fukuda, J.Itou, J.Munemori, "Pictograph Chat Communicator II: A Chat System that Embodies Cross-cultural Communication", *CSCW 2008*, Poster paper, CD-ROM, (2008)

[3] K. Dietz, and D. Leigh, "Diamond Touch : A Multi – User Touch Technology", Proc. UIST. 2001, pp. 219 – 226 (2001).

[4] Zlango http://www.zlango.com/

[5] J. Munemori, S. Ohno, T. Yoshino, "Proposal and Evaluation of Pictograph Chat for Communication", Journal of IPSJ (Information Processing Society Japan), Vol.47, No.7, pp.2071-2080 (2006).

Development of a Communications System with Pictograph Translation for Parent and Child

Satoru Fujii*, Nao Yasuda*, and Kouji Yoshida**

*Matsue College of Technology, Japan
**Shonan Institute of Technology, Japan
*fujii@matsue-ct.ac.jp
**yoshidak@info.shonan-it.ac.jp

Abstract - Children in day care facilities can experience anxiety and a sense of isolation. Equally, working parents have on-going concerns about their children's state of mind. A system to enable mail communication between children and parents has been developed. By using QR code cards and a touch screen with pictographs pre-school age students are able to create and transmit mail messages. The message is translated into text, which is transmitted to the parent. The parent's response is then converted to pictographs. Field trials and evaluations have indicated that the system is useable and has good potential. Nevertheless further improvements are possible.

Keywords: Communication, Parent and child, Translation system, Pictograph , Mobile phone

1 INTRODUCTION

The number of working mothers has been increasing over recent years [1]. This often results in pre-school age children being placed in day care facilities or nursery schools. Pressure of work may lead to parents asking the facility to keep their child for longer hours and/or parents arriving late to collect their children. This causes stress for both parent and child. The child can feel anxious or isolated and the working parent finds it hard to concentrate on their job through a combination of worry and guilt.

Therefore there is a need for mechanisms to facilitate communication between parents and young children. We selected an asynchronous communication mode to maximize convenience, especially for parents who may not be able to communicate in real time from their place of work but who can readily use a mail function.

Three to five years old children recognize pictures and animation more easily than characters. Therefore, we selected pictographs as the mechanism for exchange emails. Pictographs are already used in mail communications instead of sentences, but the pictographs used on mobile phones are not easy for children. Therefore we invented an original pictograph lexicon based on a questionnaire survey.

We developed a mail system for parent and child. Using this system, a child can create mail using pictographs and a parent can write mail using sentences. Our system translates between the pictographs and sentences. We have further developed and improved this system based on opinions from a company that was developing products for preschools. We brought our system into a preschool. Both parents and children involved in evaluation found it useful.

2 BACKGROUND

Many kinds of information equipment for children such as child-friendly mobile phones are now available on the market. However such equipment requires the user to be able to read. In addition, there are products available so that parent can see his or her child from a distance by using a camera. Mimarmorikun [2] is one such system. However, its use is limited because a parent cannot access the image via a mobile phone and it doesn't offer interactive communication. Alternatively, there is a communication system using pictographs [3], but this system is not practical for use in communication between parent and child.

We sought to develop a system with the merit of enabling communication at any time by using a mail function. Since the primary criterion was that any system must be useable by young children, we chose to use pictures and animation rather than conventional text. Prior to creating a working system, we collected questionnaire responses from 56 parents and 29 caregivers and teachers at three preschools. Our purpose was to determine the perceived value of the system and to collect information on the likely content of parent child communications.

The survey elicited the following responses. On the plus side respondents believed that parents could resolve feelings of unease about their children and children themselves might be less anxious.

On the downside, there was concern expressed about the increased burden the system may place on caregivers and pre-school teachers. The questionnaire responses found more positives than negatives in using the system so the

approach was seen as valuable.

Nevertheless, the perceived weakness would have to be resolved. So, it was clear that the interface would have to be very child friendly to enable children to use it unassisted.

Three to five years old are better at understanding pictures and animation than recognizing characters or reading/writing sentences. Therefore, the child interface uses animation. In addition, the pictures and animation are activated by touch panel rather than keyboard and mouse.

We used responses from questionnaires from 67 parents and their child to identify the most useful range of pictographs for communication between parent and child. This resulted in a set of 70 keywords, each with an associated pictograph [4][5].

3 SYSTEM CHARACTERISTICS

The communication system has a translation function between pictograph and text message. The system translates both the child's pictograph message to text message, and parent's text message into a pictograph message. A parent can use either a mobile terminal or a personal computer, while the child uses a special terminal located in their pre-school or nursery school.

Figure 1 shows the catalogue of pictographs and the screen for creating pictograph sentences. Pictographs can be selected by touching this screen, and this builds a pictograph sentence on the upper left-hand side of the screen. In this case, the pictographs are "Sad", "Greet me", "Soon", so it means "I am sad. Please greet me soon".

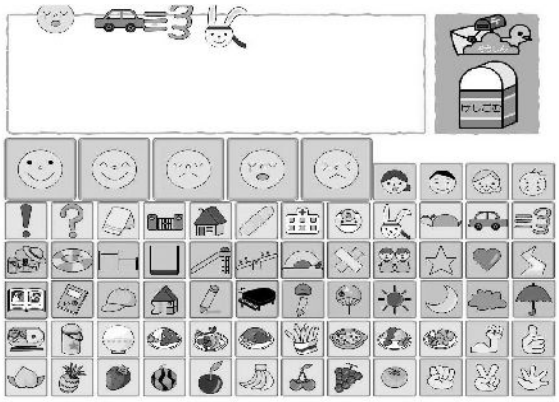


Figure 1 Screen for creating mail using pictographs

Since the system can be used with either a personal computer or a mobile phone, a parent can access it regardless of time and place. The pictograph message is translated into text and sent to the parent's mobile phone or personal computer. Conversely, the parent's text response is translated into pictographs and transmitted to a special terminal in the preschool [5].

4 SUMMARY OF SYSTEM STRUCTURE

Figure 2 shows the system structure. Children can use a terminal in their preschool. They operate it by using the touch panel of a 19 inch liquid crystal display without mouse or keyboard. They can easily log on to this terminal using the QR code reader. The children's terminal has 70 different pictographs arranged in the pattern shown in Figure 1. The server manages necessary information such as parent's address and child's name using an address list. The translation block translates between pictograph messages and text messages. The parent utilizes either a mobile phone or a personal computer. In addition, the parent can review the picture diary described in the next section using Internet.

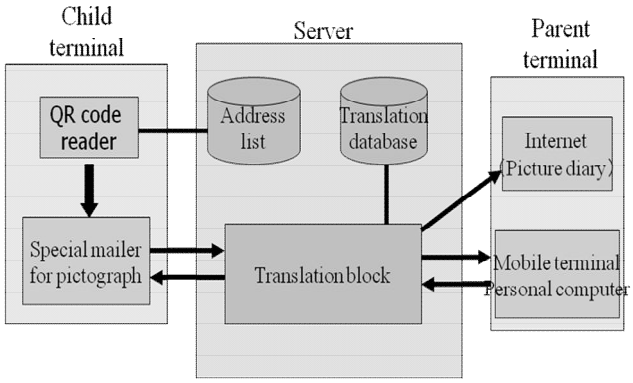


Figure 2 System structure

5 SYSTEM FUNCTIONS

5.1 Log on function using QR code

Figure 3 Nameplate with QR code

The pre-schooler logs on using a name card with a QR code shown. These are personalized for easy recognition (Figure 3). The log in screen uses animation (Figure 4) to make the process easier for small children. The QR code automatically identifies the child and the parent's mail address. Once logged on, the child can then proceed to create and send messages.

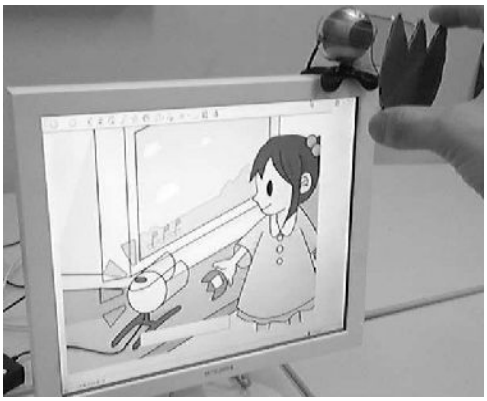


Figure 4 Log on screen

5.2 Operation using the touch panel

The child's terminal in preschool can be operated by finger touch on the panel. No mouse or keyboard is required so a child requires little if any assistance from the staff.

5.3 Pictograph translation function

A pictograph translation function in the server (Figure 2) converts pictographs to sentences and vice versa. A survey of 67 parents and 67 children enabled us to identify the key elements of conversation and led to the choice of 70 pictographs.

5.4 Picture diary function

Figure 5 Picture diary

Children's messages are stored on the server in both pictograph and sentence form. This diary is accessible via the Internet. This function enables parents and their children to review the communications later and discuss them together. We believe that this is a positive resource for parent and child communication. It also enables parents to keep a record of the messages and see patterns in the type of messages their child has sent.

5.5 Function for pictograph browsing using mobile phone

Parent can read transferred text message via mobile phone or personal computer. The pictographs her selected by the child can also be viewed. Sample screens are shown in Figures 6 (A) and (B).

(A) Transferred mail (B) After connection to URL

Figure 6 Sample screens of browsing pictograph on mobile phone

6 IMPLEMENTATION

6.1 Language

We developed the special mailer program using FLASH. Thus, we could make dynamic contents including animation , and could develop a system a child would find interesting. The program for text processing, translation and mail processing was developed using PERL.

6.2 Translation Algorithm

The pictographs are divided into five categories; "feelings", "person", "adjective or adverb", "verb and noun",

and "other". The translation program translates the pictograms in each category sequentially.

The translation algorithm is shown in Figure 7. First, the words of "feeling" and subject must be translated because the child primarily wants to explain his or her feelings to the recipient. For example Figure 7 shows that the algorithm translates the pictographs of "sad" and "mother" to "I am sad, mother" and conveys his or her feelings first. Next, the algorithm converts what the child did or wants to do by combining "adjective or adverb" and "verb and noun". As shown, if there are pictographs of "with friends" and "seesaw", the translator interprets this as "I played seesaw with my friends "This is because, to a child, "seesaw" means the activity such as "I played seesaw". On the other hand, plural pictographs are translated as "much". In this example, curry is translated " I want to eat (ate) much curry". Janken such as "Goo" or "Choki" are assigned to the "other" category and are translated last in the text message.

Order of translation: feeling→person→adjective or adverb →verb and noun→other

"Mother, I am sad. I played seesaw with my friend. I want to eat (ate) much curry rice."

Figure 7 Example of pictograph translation

7 METHODOLOGY AND OUTCOMES

We evaluated this system twice in preschools in Matsue city. On the first occasion, five year olds operated the system offline. The second time, parents and children used mobile phones and the special terminal online. Figure 8(A) is a picture from the first evaluation and Figure 8(B) shows the second trial. The evaluations were based on two items. First, "Could a child use the system unassisted? (Observations of 30 adults). Second "Did translation result agree to the child's intention?"(Responses from seven children).

We explained the operation method to groups of three children at a time. Then they each had a chance to try it out. All the children tried twice.

The children were then asked to send a message to their mothers. After sending the message, the children were interviewed regarding the content and intent of their messages.

(A) Children using the special terminal

(B) Parents and children using the system

Figure 8 System in Use

The outcomes were positive. The child participants had no difficulty understanding how to log on using a nameplate with a QR code or how to send mail. They could also use the eraser function and understand its purpose.

Six of seven children said the translation result agreed with their original message. The exception was a child who selected too many pictographs and this obscured the text translation.

The reaction of the children during field trials and evaluations was very positive. They were very enthusiastic and wanted to try it by themselves. We believe that this was in part due to the appeal of the animation and pictographs, which the children understood easily. All the children could use the system immediately, confirming that the system has potential as a communication medium.

In the second experiment, eleven pairs of parent and child used and evaluated this system. We showed standard sentences to parents, and they then created and sent messages. Each child responded to the parent's message and the parent then sent a response. After the field trial, parents answered a questionnaire. The questions were divided into evaluation of the mobile terminal functions and evaluation of the system as an effective means of communication. The questions and results are as shown

below.

(a) Function of pictograph view screen of mobile phone

Q1 How useable did you find it?

Q2 Could you see the screen and pictographs easily?

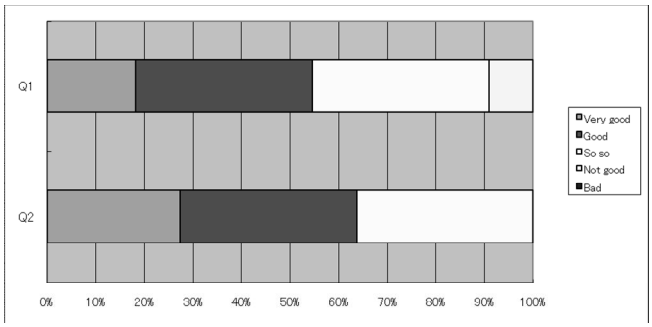


Figure 9 Evaluation of the function of mobile terminal

(b) The system as a hole

Q1 Were you satisfied with the translation function?

Q2 Did you feel this system would be effective for relieving your child's sense of isolation?

Q3 Did you think this system would be useful as a tool to communicate with your child?

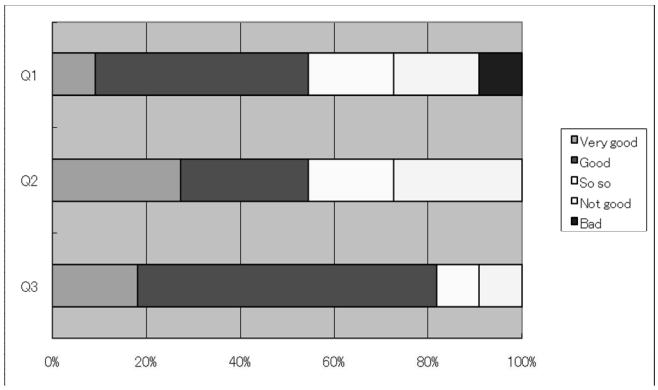


Figure 10 Evaluation of the system

As shown in Figure 9 the evaluation of usability of the mobile terminal was good. Prior to the trial, an earlier questionnaire had indicated that about 50% of parents were concerned about the additional burden the system would place on the nursery school staff. However, after the field trial, 80 % of parents agreed that this system was useful for communication (Q3 of Figure 10). Parents commented that this system was interesting and they were happy to receive mail from their children while at work. They commented that this system would be useful.

Although the questionnaire results, comments and our observations were generally positive, there are still some problems to be resolved.

We could observe as each child-parent pair communicated through mail exchange. Most children could not respond adequately to their parents, because the children had only pictograph translations of the sentences and so could not always understand their parent's intention. Whereas parents could see both pictograph and sentences. To resolve this problem, we are considering ways to also show sentences to those children who can read, and for the system to read the sentence by speech synthesizer for those who can't.

Another problem is that the pictographs cannot distinguish between past and future. For instance even if the child's intention was to convey "I ate", the translation result is "I ate or I will eat", so the message to the parent is ambiguous.

There was a problem with the pictograph table as children took a long time to select from the many pictographs to make a message. We have therefore modified the screen. Figure 11 with the new arrangement of the pictograph table can be compared with Figure 1. The expressions and foods are now ordered as shown. In addition, some pictographs that were not used by the children have been deleted. Future evaluations will include comparison of the original and new pictograph tables.

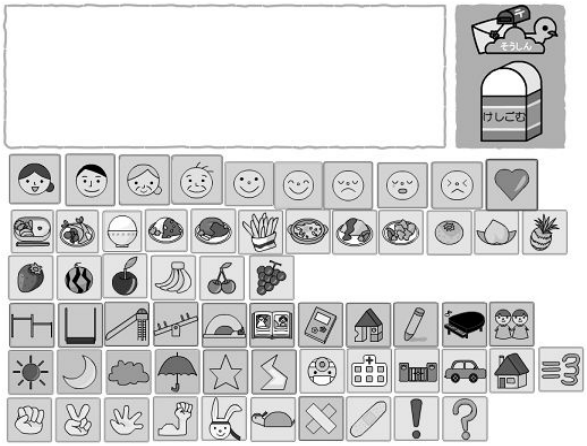


Figure 11 Example of improved screen for pictograph mail

Japanese is written using two syllabaries, hiragana and katakana plus kanji. Conventionally, children learn hiragana first followed by katakana, then, finally kanji are introduced. Many children have learnt to read words, if not sentences, in hiragana and katakana before they enter elementary school. Therefore, storybooks for younger learners are presented in a mixture of hiragana and katakana. Parents with slightly older children, five years old and above, commented that it would be convenient to enhance the pictograph message with hiragana and katakana. If the child can read a little, he or she could then communicate more smoothly. To accommodate the wider age group we are planning to offer a choice of pictographs and characters.

8 CONCLUSIONS

We developed the communication system to offer asynchoronized communication between parent and child. The system relies on a QR code log-on card that identifies the child and the target recipient. Children create messages by selecting pictographs that are translated into text and transmitted to the parents' mobile phone or personal computer. The reverse process enables parents to send messages back.

We improved the system after receiving comments from the company that is developing some products for preschools and receiving questionnaire responses from pre-school institution staff and parents. After trials that demonstrated that system use was within the capability of pre-schoolers, we ran evaluations and sought further feedback from parents and pre-school staff. This has led to further improvements such as modifying the pictograph catalogue. Additional modifications will include including an option for older children to select Japanese characters and refinements of the pictographs to include verb tenses. Enabling the child to select a recipient from father, mother or grandparent will also make it more useful. This system received the Grand Prix Award at the Chugoku Campus Venture Grand Prix in 2009.

REFERENCES

[1] http://www.gender.go.jp/pamphlet/pamphret-main/index.html

[2] http://www.kouboum.co.jp/kiji.htm

[3] Jun Munemori, Sumika Ohno, Takashi Yoshino: Proposal and Evaluation of Pictograph Chat for Communication, Jouanal of Information Processing Society of Japan, Vol.47, No.7(2006)

[4] Nao Yasuda, Megumi Shirane, Wataru Aoto, Tsubasa Teramoto, Kayoko Hayashibara, Satoru Fujii: Development of a Communication System using Pictograph Translation by Parent and Child, Chugoku chapter congress of Japanese Society for Information System in Education, No.22, pp.17-20 (2008)

[5] Satoru Fujii, Nao Yasuda, Megumi Shirane, Wataru Aoto, Tsubasa Teramoto, Kayoko Hayashibara: Development of a Communication System using Pictograph Translation for Parent and Child, Papers of Groupware and Network Service, pp.67-72 (2008)

Simulated Collaboration to Understand Japanese Offshore Software Development in China

Takaya Yuizono* and Lihua Xuan*

*School of Knowledge Science, Japan Advanced Institute of Science and Technology, Japan
yuizono@jaist.ac.jp

Abstract – Two primary issues in Japanese offshore software development in China are how to manage specifications and intercommunication. To consider the software development using groupware technology in a laboratory setting, we proposed a simulated collaboration task. In the experiment, the Chinese, Japanese, and Japanese-Chinese groups corresponded with offshore development, domestic development, and cooperative development, which is a proposed future style, respectively. The laboratory experiments' results indicated that a well-collaborated team produced a good model chart; the Chinese participants experienced difficulty in using Japanese-language communication, making some of them self-assertive; and, in comparison to the Japanese participants, the Chinese participants tended to be satisfied with their results of the model chart, considering them neither good nor bad.

Keywords: offshore software development, software specification, laboratory simulation

1 INTRODUCTION

Nowadays, both Japan and China widely experience interactions in their trade and between their people. In trade, software development is representative of work that necessitates knowledge workers, making it an important collaborative undertaking. Japanese IT offshoring came into prominence in the 1980s, motivated by pressures to reduce labor costs and, to date, has been carried out primarily in China. Offshoring in recent years requires the management of the complete process of software development [1].

The largest amount of offshoring from Japan is sent priarily to China; hence, China receives the highest number of outsourcing jobs from Japan [2]. Offshoring from Japan to China is still prominent in the software development industry, although offshoring to India and Vietnam has also increased. As is well known, all offshoring projects were not always successful because of differences in language, culture, corporate climate, and work environment.

Groupware technologies, which support distributed software development, have the potential for IT offshoring because the major issues in offshore development involve software specification and human communication. In order to understand the software development process in different cultural settings, the time taken, expenses incurred, and detailed observations or logs maintained by an IT vendor are necessary [3]. In addition, evaluating a new type of offshoring may not be realistically acceptable to the vendor. Therefore, we design a collaboration task that simulates offshoring development and is availabile in a laboratory setting.

In the following section, we describe the collaboration model for offshore software development. In the third section, we explain the experiment and proposed collaboration task. In the fourth section, we present the results of the experiment and discuss them. In the last section, we summarize this paper and suggest a future direction.

2 SIMULATED COLLABORATION FOR OFFSHORE SOFTWARE DEVELOPMENT

2.1 Collaboration Task

We proposed a work model on the basis of the collaboration between the Japanese and Chinese in offshore software development, to investigate or explore the issues that exist in this collaboration.

The questionnaire on offshore software development was based on that of Nakahara and Fujino [4]. The results showed that, in the case of Japanese companies, about 20 percent indicated a communication problem, about 30 percent indicated problems with confirmation and modification of specifications, and about 20 percent indicated a communication problem in the case of offshore companies.

On the other hand, software development based on a specification is difficult to replicate in a laboratory experiment because the work is limited to only a few hours and the recruitment of programmers for the software development is unrealistic, since many students have yet to learn the essentials of programming. Therefore, we considered model charting as the collaborative work instead of program development. The model that simulates the collaboration in offshore software development is depicted in Figure 1.

In the collaboration task, the client and vendor are located at separate remote sites. The client sends a software specification to the vendor, and the vendor develops a model chart instead of developing a software program according to the specification.

The vendor consists of three persons—the team leader and two workers, who are not programmers—and they work in the same place. In the laboratory, the workers can pose any questions regarding the specification to the leader. When the leader is faced with a question that has not been resolved, he/she can question the client, who is in a different location.

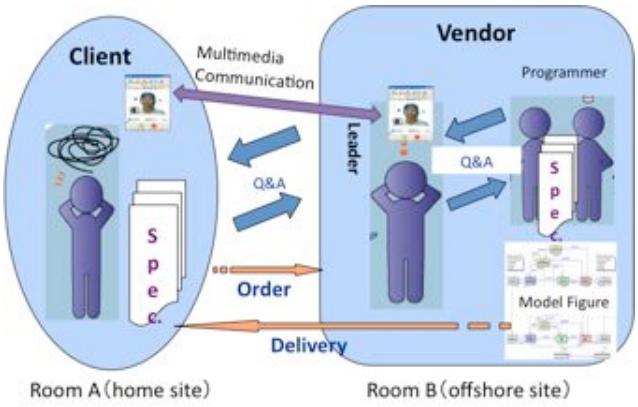


Figure 1: Collaboration model of offshore software development

The vendor prepares a system chart on the basis of the specification that is assumed to be ordered by the client. For this collaboration, workers can use a paper and pencil, but they must depict the chart as a final groupware product that supports a shared screen and chart-making function. The workers use a label and arrow representation for the charting. The leader can check the chart on the shared screen using the groupware.

In the task, two versions of specifications were prepared in order to mimic a problem caused by a specification change, which is considered a typical cause of confusion in software development [4]. The modified specification is an advanced version of the specification that was initially ordered by the client.

Three team patterns are considered because the characteristics of offshore software development are expected to be revealed in the comparison of these patterns. In the first pattern, all the people belonging to the vendor's side are Chinese. This simulates ordinary offshore software development and, hence, is labeled the "Chinese Group." In this case, the team leader is a Chinese who is fluent in Japanese, but the workers cannot speak Japanese though they can read Japanese sentences. In the second pattern, all the people are Japanese and are labeled the "Japanese Group," in order to simulate a domestic software development team for the comparison. The third pattern is considered the future style of offshore software development and is reflected by a collaboration of Japanese and Chinese participants and, hence, is labeled the "JC Group." In this pattern, the leader is Japanese and the workers are Chinese.

2.2 Software Specification for the Collaboration Task

The software specification was prepared before the collaboration task. The content of the specification was taken from a standard problem on stock management, for software specification research [5].

The theme for problem solving in the collaboration experiment was entitled "The store management system for a liquor company" and was developed to compare program design methods.

The structure of the specification was referred to as the guideline for the requirement specification [6], and the language description in Japanese follows Ooki et al's explanation [7]. The Japanese description reflects concurrent object processing and has no symbolic description that depends on a specific program design technique. Nowadays, a description with these characteristics suits the specification description for modem object-oriented development. A part of the specification is depicted in Figure 2.

In the specification, there are descriptions of the aim of the systems, glossaries, requirements, and specifications. In the first section, entitled "Introduction," the basic knowledge of the system is explained. In the next section, which is a subsection, the behaviors are explained according to the object. The objects named "Receptionist for stock," "Receptionist for delivery of goods," "Storage," and "Deficiency of stocks" are defined; accordingly, a behavior that sends an event to each object is defined.

The parts of the sentence that were underlined were added when modifying the specification during the collaborative work to simulate a specification change, which often occurs in software development and causes difficulties in the development.

1. Introduction
1.1 Goal: Model of Stock Management System
1.2 Scope: Management System of Liquor Shop
1.3 Glossary
Stored object: Commercial product in storage corresponding to each product.
Container: A container loaded with mixed products.
List of stored products: One item from the stored list and memorized correspondence between the stored product and container.
2. Requirements and Specification
2.1 Receptionist for the stock
2.1.1 When a container arrives
2.1.1.1 Making a new container
2.1.1.2 According to each branded product
a. Update the number of stocks to storage
b. Making an item list of the cargo.
2.2 Receptionist for the delivery of goods
2.2.1 When it receives a request for the delivery of goods
2.3 Storage
2.3.1 When it accepts the number of stocks from the receptionist.
2.3.1.1 The number of stocks become "the number of stocks + the number of entry stocks" and update the number to a deficiency of stocks.
2.3.1.2 When the delivery of goods necessarily occurrs because the deficiency of stocks is dissolved.
a. ~
b. ~
c. ~
2.4 Deficiency of stocks
~

Figure 2: Part of the specification with a stored system (underlined specification was added during the collaboration as a modification)

3 EXPERIMENTS WITH THE COLLABORATION TASK

3.1 Experimental Procedure

There were 18 participants recruited for this research—eight Japanese and ten Chinese—and they were organized into six groups. All the groups assumed the role of vendors and each comprised one leader and two workers. A Japanese teacher assumed the role of the client in all the six experiments. In the case of the "Chinese group," all the members were Chinese; in the "Japanese group," all the members were Japanese; and in the "JC group," the leader was Japanese while the two workers were Chinese.

The procedure for the experiment was as follows. In the beginning, for about 15 minutes, the experimenter explained the collaboration task and how they should use the system. She described the specification and a chart of the system from a textbook on software engineering [8]. In the chart, a label object is used to represent an object in the specification and an arrow is used to represent the sending of an event; the arrow is often added to the string using a label object to explain the event. Then, the participants began the task. The total time for the task was fifty minutes. This time was set from the results of two experimenters, who completed the task in fifty minutes, using computers. At the beginning of the task, the participants were informed that the total time for the task was thirty minutes, and that it should be delivered with the specification of the storage system. The A3-sized paper and the pencil were provided for free usage at the same time. When it was twenty minutes into the task, the experimenter, who was the client, informed the participants of a modification to the specification and delivered the modified specification to them.

After completing the task, the participants answered a questionnaire that was based on a five-point scale and checked the relevant difficulty level with regard to their understanding of the specification. If they checked a low value, they were prompted to write a reason for this. Moreover, they were urged to underline the parts they did not understand and provide reasons or opinions for the same.

3.2 Environment

The experiments were carried out in two rooms—the faculty room and the research staff room—at the Graduate School of the Japan Advanced Institute of Science and Technology. Skype, a well-known software application that allows voice calls over the Internet, was utilized to communicate between the client and the leaders at the vendor site, primarily for questions and answers. The vendor used a groupware called KUSANAGI [9] to create a system chart. The groupware had a brainstorming tool, grouping tool, and arrow tool to support the grouping stage of the distributed and cooperative KJ Method [10]. A picture of the experimental setting at the vendor site is depicted in Figure 3. A screenshot of a modeling chart is shown in the next section.

In order to create the chart using KUSANAGI, the user labeled the objects or event explanations and depicted the event flow using arrows.

Figure 3: The experimental setting at the vendor site: On the left are the two workers and on the right is the team leader.

4 RESULTS AND DISCUSSION

4.1 Results of the Collaboration

The six charts of the storage system referred to the specification that was obtained from the experiment. The evaluation points of the charts and the knowledge background along with which team prepared the charts are summarized in Table 1.

Table 1: Evaluation of model charts and background of the group

Group name	Chinese Group-A	Chinese Group-B	Japanese Group-A	Japanese Group-B	JC Group-A	JC Group-B
Specification score (correct percentage)	12 (37.5%)	23 (71.9%)	17 (53.1%)	21 (65.6%)	4 (12.5%)	20 (62.5%)
Number of labels (correct percentage)	14 (78.6%)	24 (95.8%)	24 (83.3%)	27 (92.6%)	17 (35.3%)	26 (80.8%)
Number of arrows (correct percentage)	8 (100%)	21 (100%)	21 (66.7%)	20 (95.0%)	11 (36.4%)	27 (77.8%)
Knowledge of IT	1	3 (All)	3 (All)	3 (All)	3 (All)	2
Learning experience of the specification	1.7	4.0	3.7	2.7	1.7	1.7

We evaluated the charts corresponding to the specification. The specification described in Section 2 has thirty-two lines. We checked the reflection of the contents on each line of the specification. We marked a circle when the line was adequately reflected in the chart, a triangle when the line was reflected to some degree in the chart, and a cross when the line was not reflected in the chart. Two persons performed this evaluation: one was a client and the other, an experimenter. We assigned 1 point to a circle-marked line, 0.5 points to a triangle-marked line, and 0 points to a cross-marked line. The total number of points from all lines implied the evaluation value of the chart. We named this value the "specification score." The marking between two evaluators indicated a high correlation coefficient: 0.75 from each line and 0.99 from each chart.

In addition, we presented the number of labels and arrows in each chart in Table 1. As described in Section 3, the labels represented an object or an event, and the arrows represented an event flow. We checked the correctness of the labels and the arrows by referring to the specification. The procedure for the check is similar to the procedure for the specification score. The score for the evaluation value of labels and that for the evaluation value of arrows are calculated as correct percentage on the table 1.

The knowledge background of the group was self-reported, indicated by the number of persons who answered "yes" to knowledge on information technology, and the average score from the five-scale questionnaire on software specification. When the value is five, it implies that the person has learned well, and when the value is one, it implies that the person did not learn at all.

In the results, the Chinese Group-B scored about seventy percent and the Japanese Group-B and the JC Group-B got more then sixty percent. The Japanese Group-A continuously scored about fifty percent. The Chinese Group-A scored about forty percent, and the JC Group-A, only about ten percent. The charts of the groups that scored more than sixty percent had many arrows and many labels compared to the other charts. The charts by the Chinese Group-A, which had a very low score, had fewer labels and fewer arrows, and the correctness of representation was inferior. The model chart prepared by the Chinese Group-B, which had a high score, is depicted in Figure 4, and the model chart by the JC Group-A, which had a low score, is depicted in Figure 5.

Before the experiment, we assumed that a knowledge background affects chart making. However, there are no such characteristics with regard to the knowledge background in the results. The experience with software technology did not always lead to good results. This may imply that the proposed task did not require software technology skills such as programming.

4.2 Results of the Questionnaire

The results of the questionnaires taken after the collaboration task are described below, and the results with the five-scale evaluation are shown in Table 2. A 5-point evaluation implied very good and 1 point implied very bad. From the questions, Q.8 and Q.9 were given to only the workers. We added a star to the left of the table in the case that showed a significant difference in the ANOVA-analysis between the value of the Japanese and Chinese answers.

Table 2 illustrates the interest level of the participants in the work task, levels of cooperation, communication within their groups, and ability to clearly pose questions to the client.

From the perspective of differences between the Japanese and Chinese, the Chinese tended to be more satisfied with their system charts than the Japanese. The Chinese, who belonged to the Chinese Group-A and the JC Group-A, which had low scores, responded that they were satisfied with their results. These results could lead to confusions in software development, so it is necessary to periodically check if the progress of the development is sound.

According to the questions to the workers, the Japanese participants communicated well with the leaders of their groups, but the Chinese participants felt that their communication was neither good nor bad. The Chinese colleagues could speak their mother tongue among themselves, but they were required to communicate in the Japanese language in the case of the JC Group, which is assumed to be the future style of offshore development. In other cases such as questions to a client, the Japanese participants felt that their communication was good, but the Chinese were indifferent about theirs. It is assumed that the differences in language clearly affected the conscious effort to create cooperative communication.

Table 2: Results of the five-scale questionnaire

Items	Value
Q1: Understandability of the collaboration task	3.3
Q2: Interest in the task	3.8
Q3: Pre-image of the system modeling	3.1
Q4: Understandability of specification	2.6
Q5: Satisfaction with the chart	2.5*
Q6: Is the work collaborative?	3.7
Q7: Do you communicate within a group?	3.6
Q8: Can you question your leader?	4.6*
Q9: Does your leader understand your question?	4.4*
Q10: Do you communicate well with the client?	3.3*
Q11: Can you pose a question to the client?	4.1
Q12: Does the client understand your question?	3.7*
Q13 Do you feel the barriers imposed by different cultures?	2.6

*Significant difference between the Japanese and Chinese with the ANOVA-test, $p < 0.05$.

Doubts on the specification were highlighted; after the analysis of the experiment, it was found that twenty out of the thirty-two lines were questioned by the participants owing to their lack of comprehension. The common doubts posed by both the Japanese and Chinese participants comprised five lines that included the repetition of words such as "container" or the "deficiency of stocks," phrases that were ambiguous such as "because the deficiency of stocks is dissolved" or "when a delivery of goods necessarily occurred," and technical terms such as "chain." In addition, for the Chinese, their characteristic doubts

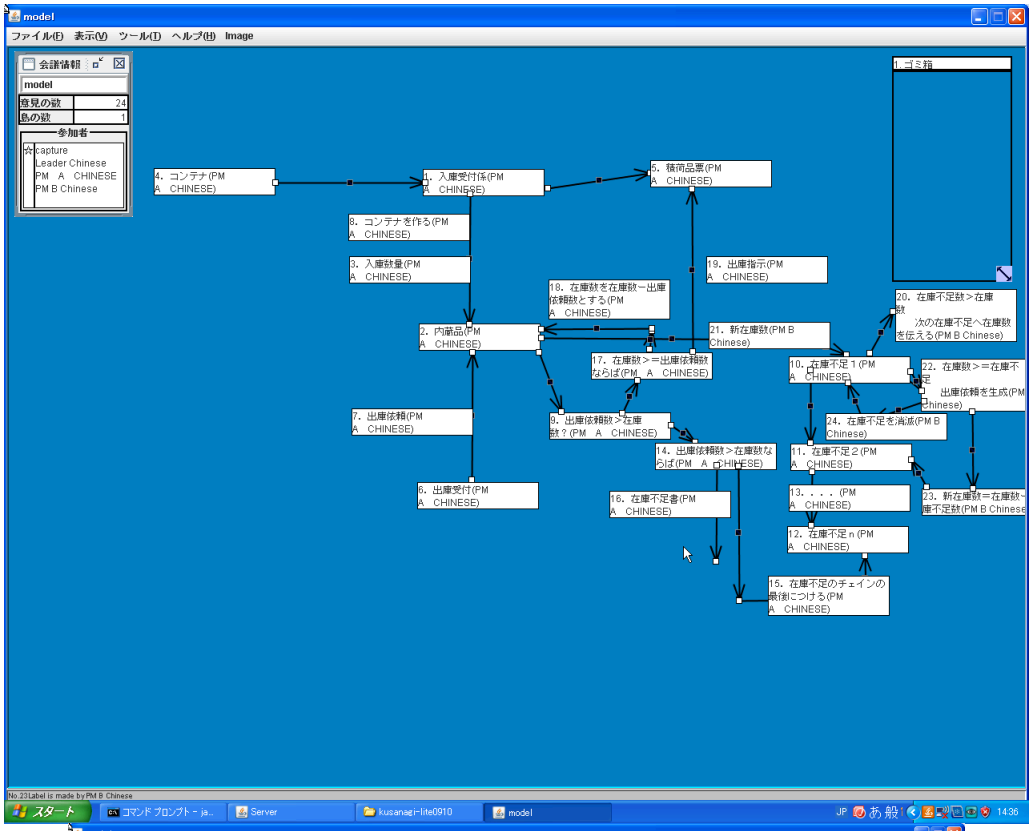


Figure 4: Screenshot of a Good Model Chart created by the Chinese Group-B

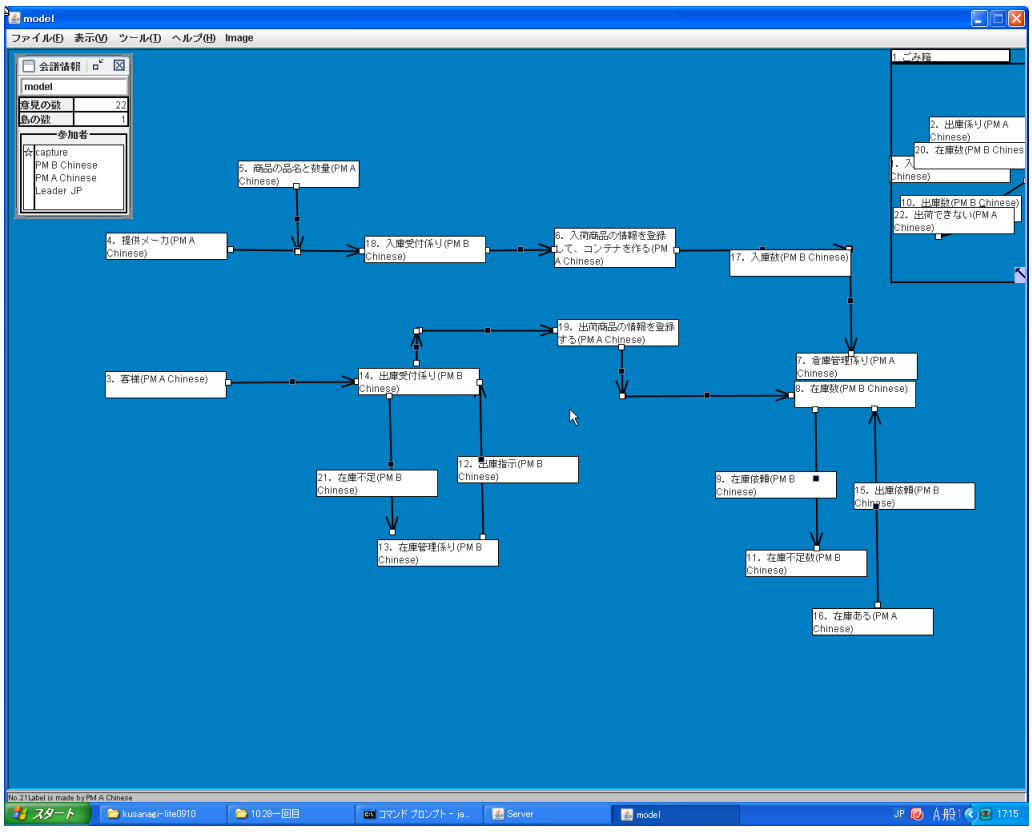


Figure 5: Screenshot of a Poor Model Chart created by the JC Group-A

Table 3: Video observation of the collaboration.

Group	States of collaboration	States of communication
Chinese Group-A	The two workers participated individually in the work. One worked on the paper and then on the system. The other worked directly on the system.	Opinions were exchanged between the three persons. One worker was self-assertive.
Chinese Group-B	The two workers did not divide the work. One worked on the paper and the other on the system, on the basis of the results of the paper.	The three people collaboratively worked on each line of the specification. They spontaneously exchanged their opinions.
Japanese Group-A	Neither worker was given a share of the work. The leader controlled the work sharing, and the workers only worked on the system.	There were a few sets of good communication between the three participants. The leader identified the orders to the workers for each line of the specification.
Japanese Group-B	The two workers were each assigned a share of the work. One worked on the paper and the other on the system, on the basis of the results from the paper.	One worker was silent while the other actively and successfully communicated with the leader.
JC Group-A	Each of the two workers handled a share of the work. The two first worked on the paper and then on the system.	There were exchanges of opinion between the leader and workers, but only a few between the workers. One worker was self-assertive, while the other asserted negative words.
JC Group-B	Each of the two workers was assigned a share of the work. The two directly worked on the system.	There was communication between the three participants. One worker was self-assertive, and the other asserted less qualitative opinions.

stemmed from the definition of words, such as “container,” “stored object,” and “deficiency of stocks”; ambiguous representations, such as the “next deficiency of stocks” and “oneself”; and the obscurity of outputs, such as “output a document about deficiency of stocks” and “create a request for delivery of goods.”

The abovementioned Chinese participants had a disadvantage because of the use of Japanese, which is not their mother tongue. The Japanese participants should pay attention to clear communication in the Japanese language to ensure good collaboration.

4.3 Observation of Collaboration

We described a state of collaboration in Table 3 by using a video analysis of the participant’s cooperativeness and communication.

Drawing attention to the Chinese workers, the self-assertiveness of some of them stands out. This might reflect a cultural habit that the Chinese are more individualistic than the Japanese, who prefer homogeneity. Naturally, the cooperative Chinese group created the good chart.

The Chinese Group-A, JC Group-A, and JC Group-B had less learning experience with software specification. From these groups, the JC Group-B, who created a good chart that earned more than sixty percent, worked by having good communication between the three members and by sharing the work on the computer screen. On the other hand, the Chinese Group-A and JC Group-A had to balance the work by creating parts of the chart individually first and then combining them.

Using these observations, the group that scored the lowest points had smaller amounts of work sharing and some of them were self-assertive; that is to say, they did not collaborate in the task. This implies that good collaboration favors the success of a proposed task.

In this experiment, we allowed individual work on paper but urged the usage of the shared environment and groupware technology to encourage shared work. This can be explored by combining the consideration of effectiveness and shared work in software development.

5 CONCLUSION

In this paper, we design a laboratory experiment to simulate offshore software development between the Japanese and Chinese. We performed an experiment in which the participants prepared a system chart instead of program development, using the prepared specification. We considered three types of collaborations, such as an offshore type, domestic type, and a more collaborative type.

The results of these experiments were as follows:

(1) The most collaborative team produced a good model chart.

(2) The Chinese participants faced difficulty in using the Japanese language to communicate, and some of them were self-assertive.

(3) The Chinese participants, compared to the Japanese, tended to be satisfied with their results of the model chart, considering them neither good nor bad.

In the future, we will consider an interface that elicits a collaborative mind or considers the cultural habitat. In addition, more investigations on collaborations in the software development process will be issued.

REFERENCES

[1] H. Tuji, T. Moriyasu, and T. Mori, Evolution of Offshore Software Development and Tacit Knowledge of Engineering, IPSJ Magazine, Vol.49, No.5, pp.551-557 (2008) (in Japanese).

[2] MIC in Japan, Report of Progress of Offshoring and its Effect (2007) (in Japanese).

[3] D. Perry, N. Staudenmayer, and L. Votta, Jr., Understanding and Improving Time Usage in Software Development, In Trends in Software. Vol.5, Software Process, A. Wolf and A. Fuggetta, Eds. John Wiley and Sons, New York (1995).

[4] T. Nakahara and H. Fujino, The Application of UML in the Offshore Project, Journal of the Society of Project Management, Vol.10, No.6, pp.9-14 (2008) (in Japanese).

[5] T. Yamazaki, Explanation of the Programming Design Method with the Open Problem, IPSJ Magazine, Vol.25, No.9, pp.934 (1984) (in Japanese).

[6] Japan Users Association of Information Systems, A Guideline to Requirement Specification (UVC) (2007) (in Japanese).

[7] Y. Ooki et al., Description of Stock Management System by Concurrent Prolog, IPSJ Magazine, Vol. 26, No.5, pp.470-476 (1985) (in Japanese).

[8] T. Nakadokoro, Software Engineering 2nd Edit, pp.42-46, Asakura-syoten (2004) (in Japanese).

[9] T. Yuizono and J. Munemori, Groupware for a Knowledge Creative Process of a Research Group and its Application to Externalization and Combination Steps, Journal of IPSJ, Vol.48, No.1, pp.30-42 (2007).

[10] J. Munemori and Y. Nagasawa, GUNGEN: groupware for a new idea generation support system, Inf. and Soft. Technol., Vol. 38, No. 3, pp. 213-220 (1996).

ACKNOWLEDGEMENTS

This research was partially supported by The Ministry of Education, Culture, Sports, Science and Technology (MEXT) and the Grant-in-Aid for Young Scientists (B) 21700133, 2010.

Trial of a distance learning system using a brain wave sensor

Kouji Yoshida†, Shin Maruyama†, Isao Miyazi††, Kunihiro Yamada †††, Satoru Fujii ††††

†Department of Information Science, Shonan Institute of Technology, Japan
††Faculty of Informatics, Okayama University of Science, Japan
†††Professional Graduate School of Embedded Technology, Tokai University, Japan
††††Matsue National College of Technology
yoshidak@info.shonan-it.ac.jp, 07a3131@sit.shonan-it.ac.jp,miyaji@mis.ous.ac.jp,
yamadaku@tokai.ac.jp, fujii@matsue-ct.ac.jp

Abstract—We developed a self-study system using Perl language. To date, no device has been available to observe a human state effectively. However, brain wave sensors have become inexpensive recently, allowing observation of the human state. We analyzed the information effectively; a brain wave sensor for human feedback has become usable.

This feedback system can measure a student's state of concentration and a state of a motivation. Additionally, students show raised motivation because the system transmits a message indicating motivation. Sensing that motivation, effective methods can be recognized. This report presents our results.

Keywords: Brain wave sensor, Perl language, Distance Learning, e-Learning, Blended Learning, e-Collaboration

1 INTRODUCTION

In recent years, growing interest has arisen in the progress of information and communication systems, high-speed networking, and multimedia environments[1]. Software development has become large-scale and complicated. Consequently, independent systems have become increasingly rare. Student skills related to system design and communication are therefore not good. We must consider learning systems that exploit group communication and iteration of practice to develop good quality software[2].

During cooperative software development, the same information must be taught repeatedly for it to become practical knowledge[3][4].. We produced a distance education system that can instruct students repetitively, but it is difficult to complete a program using this distance learning system alone.

To date, it has been extremely difficult to respond while grasping the state of a student using a distance learning system[5]. Nevertheless, we can cheaply use a brain wave sensor to observe the state of a student. We experimented on construction of a system for learning while using this feedback[6].

2 PROBLEMS AND PRESENT CONDITIONS

From the beginning, remote education systems have presented the problem of whether or not the learner is well accustomed to accessing the necessary media electronically. The learner is isolated: aside from the learning system itself, they can contact only an instructor. Therefore, the learner is usually apprehensive, wondering whether the system will behave as expected, whether the system is useful as expected, and whether it is possible to access the necessary contents associated with the received lectures.

It is very useful to measure the feeling of a student and their level of consciousness using distance learning systems. Such a system can feedback information from the students.

3 SYSTEM OVERVIEW

This system consists of basic software ideas for general learning based on the Perl language, including specifications for the specific learning method. In addition, the system includes an information bulletin board and chat room. It is possible to exchange messages with other learning members. The system shows the students' grade situation by ranking their relative progress.

This system is supported by both Linux and Windows operating systems. A student accesses the web browser via a personal computer. Students start advanced learning and group work using learning support and communication support. Then they start personal learning and group work with learning support and communication support. They can check their degree understanding by solving some problems for confirmation. A learner and lecturer can communicate. They can examine a function of a language using a database. An image of this system is presented in Figure 1.
In addition, the system can measure the state of learners such as concentration power using the brain wave sensor.

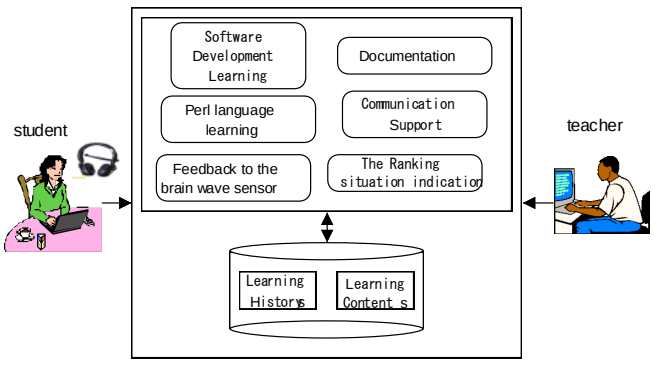


Fig. 1 . System image.

4 CONTENTS OF LEARNING

4.1 Learning Software Development

When students learn software development, they learn the contents on the Web. These contents for learning are generated using Power Point (Microsoft Corp.), but are then converted into HTML format. The contents are separated into five levels of learning for software development. Confirmation questions are prepared at each level. Moreover, if all chapters are completed, then students must answer a series of 20 questions.

・Support contents
Software development, requirement analysis, software design, program testing and maintenance

4.2 Documentation

When students are working on problems using the program, the system checks their understanding of the situation. The learning contents generated using PowerPoint are changed to HTML. The specification's documents are classifiable into three categories:

(1) External specifications form
(2) Internal specifications
(3) Test specifications

After each level is completed, a validation test is performed, followed by validation planning. Through a selection process, we correct the students' understanding. The documentation content is shown in Figure 2.

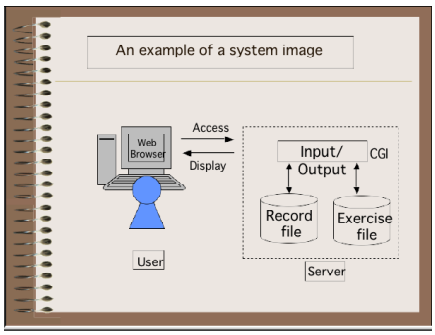


Fig. 2. Documentation.

4.3 Using Perl Language

(1) Usage of Perl language
The programs are classified according to the item and display a step-by-step process on how to use Perl language functions. Each function is then displayed entirely making it easier to examine the parameter. The main functions are printf, scanf, if-else, array, for and while.
(2) Exercise to understand the Perl program
The purpose of this exercise is to confirm what content was understood by the student from using the given Perl explanations. The exercise is selective and provides a percentage representing the degree of comprehension.
(3) Perl Program Exercise
This displays both the Perl program's mock validation exercise and the implementation section. Step 1 consists of validation, and it displays the content (e.g., the parameters and results of the program). After the students input their functions into the text field, they can confirm the entry by running the program.
When students do not comprehend the basic function, an example answer is displayed. They can confirm it through this step. This creates a simple explanation that is sufficiently clear for a novice programmer to comprehend. Step 2 consists of implementation in an exercise form. They can experience a mock implementation and debugging of a complicated program. These ideas are displayed in Figure 4.

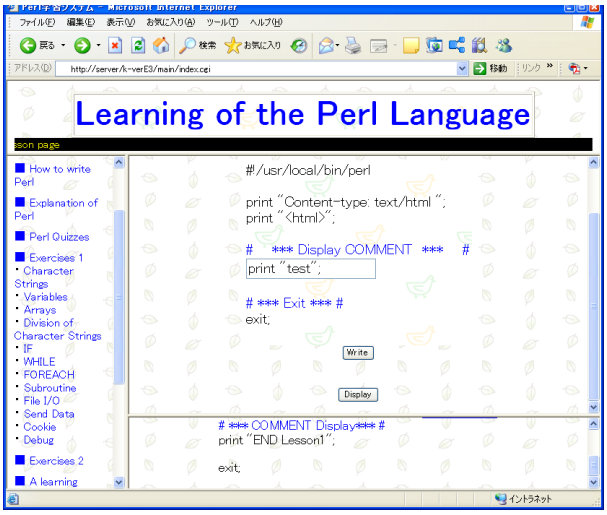


Fig. 4 Programming image.

・Discussion Board Q&A
This section is used for questions and for information exchange between the teacher and students. This section also allows the teacher to create or respond to messages. The page's background color changes whenever a new message is posted alerting students of the message.

・Message Exchange
This section carries sent questions and other information exchanges on a peer-to-peer level. Even with encrypted mail addresses, messages can be sent using a user-defined nickname.
(1) Mail transmission function and data exchange
(2) Mail receiver function
(3) Mail transmission and history function
(4) Learning schedule setting and communication function

4.4 Support for the module

This support function programming technique lets learners practice basic program-linking to enable the learner to experience test trials to learn how to link modules.
Listed below are the practice steps to be followed.

· Practice of maintaining data integrity
· Practice of correcting programs for linking
· Practice of multi-program linking as a test trial

Following are additional descriptions for practice.
(1) Practice of maintaining data integrity
This is provided to the learner to show points of caution when exchanging data among modules. Every learner is expected to join a group for this purpose; everyone in the group is expected to enter variable names or real numbers that come to mind. Through this process, among other things, the learner is expected to learn how important it is to use consistent variable names in a specification document.
(2) Practice of a correcting programs for linking
The learner practices programming for module linking by giving the learner a program for use to link some modules in which at least one error is included intentionally. Consequently, the learner must correct the erroneous portion to finish the module linking. In this practice lesson, a mode of giving special attention was provided, by which the color of the program line number changed when the learner corrected the wrong line mistakenly, or when the learner put wrong information related to a line, even though the line number itself was correct.
(3) Practice of multi-program linking as a test trial
The learner is provided with a program in which some program statement portions are intentionally left missing. The learner is then expected to complete the program to make it work properly while simultaneously reviewing and checking the associated specification. The program has been left intentionally as missing an important segment to exchange data when linking modules. Then the learner is expected to complete such an incomplete program, thereby learning the importance of data structures that are used when modules are linked.

4.5 Ranking Identification

A teacher can follow the learning progress of a student using the Web application. Students can track their progress status using a clear bar graph. The system delivers each Q&A using a mailing list and searches the mailing list's archives. From the instructor's side, it is possible to see the progress of students. This allows the instructor to measure the gap separating students. It is also possible to add explanations or hints for specific lessons, such as helpful teaching materials to assist the students' learning and progress further.

Students can also receive an explanation of the lecture via PowerPoint (Microsoft Corp.). When students wish to see a lecture's contents, they can observe the contents on their personal computer using the HTML conversion.

Each student can browse practicum contents using a personal computer. The content advances according to the student's own progress. The practicum content is created from JavaScript and is displayed as PowerPoint (Microsoft Corp.) changed to be displayed as HTML data. After the students solve the practicum problem by attending lectures, the answer is submitted on paper. Students can see their relative ranking in the class at any time. The teacher can easily supervise each student's progress, thereby allowing a direct channel to advise those who have fallen behind.

If a question or comment is brought up during a lecture, that question could be posted in the discussion board's Q&A section for all students to access easily. The background color of the page would be changed when a new message is submitted allowing for rapid notification to students. Students have access to contact the teacher and other students via e-mail.

Students can learn in their spare time while attending school for lectures and can then pose face-to-face questions. The program allows students to learn from the privacy of their home using a mobile phone or personal computer. The program allows each student to plan their own schedule. If a student falls behind schedule, then a reminder message is sent to their mobile phone. The problems presented to students in the study support section of this system are presented in a multiple-choice format. These are prepared before lectures. Students can exchange information with the teacher and other students using the communication function.

The teacher can observe any particular problem that students are working on at any time throughout the process. This function displays each student's progress in relation to the whole group. In addition, students can see the rankings of other students who are solving the same problems. The practicum's progress is displayed for students as shown in Figure 5.

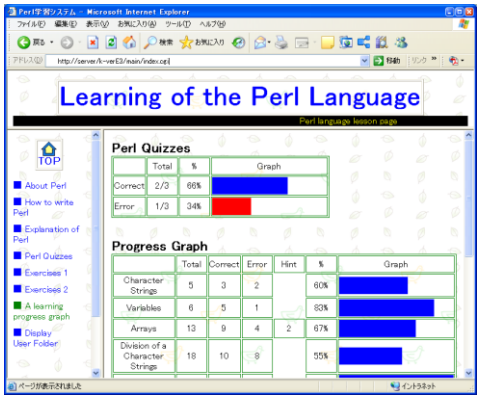


Fig. 5. Bar graph showing progress.

The teacher can supervise the class by observing the list of problems that each student is currently solving. This allows the teacher to advise any student who has fallen behind, as well as students whose progress has stopped altogether. This function also permits the teacher to track the overall progress and understanding of the class' content better.

5 RESULTS AND EVALUATION

5.1 Period and Method

We used a seminar method and a distance learning system for third-year students of our department [7]. The period

was four months. The contents that we used were shared with three groups and used as described below. We divided 10 students into three groups with 4, 4, and 3 members.

First, the installation of Apache and the Perl language are expected to be done on the local server. Thereby, the learner is expected to understand how and in what combination the Perl language would work with the Apache server as a system overall. Such a work of understanding would include a problem for which knowledge that the learner has acquired is insufficient. In other words, support from the instructor is necessary: otherwise, the problem might be difficult to resolve unless a quick response to the questions the learner might have is given through a blended learning technique.

Subsequently, premised on the understanding of c language, students learn about the relation between html and cgi. Students learn themselves through self-study about a Perl language function that is supported by Perl language. For the ensuing month, students learn the basics of Perl language to file access. Then they understand the entire Perl language. They learn basic information related to connections among programs using a program combination support function [8].

Students started group work in the third month. First, they chose a leader among the members. Next, they discussed problems among groups and decided the subject of the program they wanted to make. In these circumstances, they learned using this support system, along with communication through mail and chat facilities. Seminars were held twice a week.

5.2 Contents and Results

After they decided which program to make in their groups, they chose a title and performed basic specifications design. They decided the charge part of a program among members after having determined an external design and a user interface. Students designed the data structures of programs of the charge program. Each member wrote the documentation. After each program that they debugged was completed, they combined the programs and reached completion.

They chose "(1) A bulletin board that recorded an access history and blackjack; (2) An electronic shopping lacing braid purchasing system[The example images display by Figure 6]; (3) A Web page that combined touch-typing with a game".

Use of the brain wave sensor revealed that all groups confronted difficulties related to file input and output by CGI and a combination of programs. Programs were finished through group work. They were improved through a combination of a program to arriving at high technology through a generation process. Standardization and documentation were improved at a certain level using this practice as well.

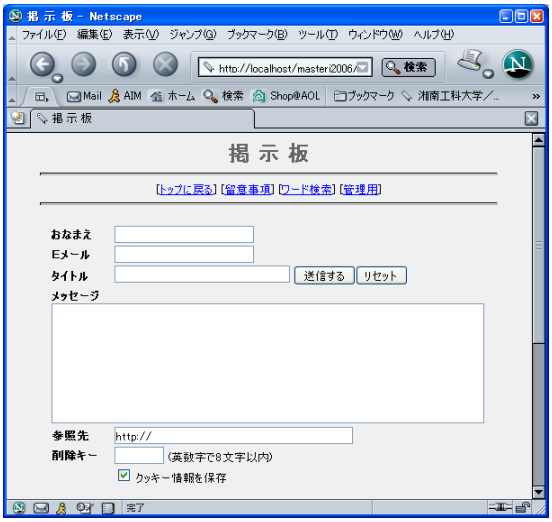


Fig. 6. Screenshot of an e-Shopping Mall purchasing system.

5.3 Results and Evaluations

Evaluations of the self-study supporting functions of Perl differed depending on the level of programming used by the student. Regarding support for combining the programming, although comprehension grew by presentation of examples showing the difficulty of combining programs and problem occurrence while remaining conscious of documentation and standardization, it remained at the knowledge level.

In the meantime, with electronic communication, although information and knowledge are conveyed somewhat, detailed explanation of complicated contents is difficult. Therefore, a means for explaining concrete images and complicated matters briefly using motion pictures and illustrations in addition to letters is necessary. Furthermore, what one has to say can only slightly involve emotional content. In particular, students are not accustomed to electronic communications. They are not good at using mailing lists and chatting.

Under these circumstances, the brain wave sensor was used; the obtained results were fed back. Then we obtained the following findings:

(1) Studying programming languages, such as Perl language learning, and repeated learning relieve stress from learning and easily sustain motivations for a long time.

(2) Inexperience in preparing specifications or the like which are complicated and hard to understand, and items that necessitate communications with others such as combinations of programs are stressful. If not done well, then a remarkable reduction in motivation arises, which is detected by the brain wave sensor.

5.4 Evaluations and Discussion

Use of the brain wave sensor enables objective measurement of motivation and concentration power of the learner. In general, this enables detection of items that learners have difficulty understanding and those which cause stress. It is also possible to know favorite and weak

subjects of individual learners, motivation depending on feelings that day, and the degree of concentration power.

Therefore, it is considered that construction of a finely designed follow-up system is possible using a distance learning system into which this brain wave sensor is built.

6 CONCLUSIONS

To date, blended learning has been incorporated into the distance learning system and effects of group learning were used. Under these circumstances, the brain wave sensor was used in this study, enabling effective measurement of the state in which the learners are learning.

As expected, results reveal that learners are in stressful circumstances in preparing specifications that are difficult to understand and in preparing combinations of programs that necessitate communication. In contrast, repeated learning and language learning such as Perl language, in which examples are presented clearly, support stress-free learning.

In the future, the authors intend to analyze functions of the distance learning system that ensure better effects and to analyze features of contents through detailed assessment of these findings and verification of their applications. This study received support from the scientific research cost subsidy "22500949" and from the organizations named above.

REFERENCES

[1] Lisa Neal, “Virtual Classrooms and Communities”, In Proc. ACM GROUP'97 Conference, Nov. 16–19, 1997, Phoenix, AZ.

[2] Natalia Lopez, Manuel Nunez, Ismael Rodriguez, and Fernando Rubio, “Including Malicious Agents into a Collaborative Learning Environment,” ITS2002 LNCS 2363, pp. 51–60, (2002).

1. Lisa Neal, “Virtual Classrooms and Communities”, In Proc. ACM GROUP'97 Conference, Nov. 16–19, 1997, Phoenix, AZ.

[3] Raymond A. Dumont, “Teaching and Learning in Cyberspace”, IEEE Transactions on Professional Communication Vol. 39, No. 4, pp. 192-204,1996.

[4] Natalia Lopez, Manuel Nunez, Ismael Rodriguez, and Fernando Rubio, “Including Malicious Agents into a Collaborative Learning Environment”, ITS2002 LNCS 2363, pp. 51-60, 2002.

[5] Kazuhiro Nakada, Tomonori Akutsu, Chris Walton, Satoru Fujii, Hiroshi Ichimura, and Kouji Yoshida, “Practice of Linux Lesson in Blended Learning”, KES'2004, Proceedings, Part 2, Knowledge-Based Intelligent Information and Engineering Systems Part 2, pp. 920-927(2004).

[6] NSW Department of Education and Training, “Blended Learning”, http://www.schools.nsw.edu.au/learning/yrk12focusareas/learntech/blended/index.php

[7] Kouji Yoshida, Isao Miyaji, Hiroshi Ichimura, and Kunihiro Yamada, “Web application construction by group work and practice”, Knowledge-Based Intelligent Information and Engineering Systems: KES 2008-WIRN2008, pp. 144-151.

[8] Kouji Yoshida, Isao Miyaji, Kunihiro Yamada, and Hiroshi Ichimura, “Distance Learning System for Programming and Software Engineering”, Knowledge-Based Intelligent Information and Engineering Systems: KES 2007-WIRN2007, pp. 460-468.

Embedding a Viewer's Face to an Advertisement for Increasing Attention

Tomoo Inoue*

*Graduate School of Library, Information and Media Studies, University of Tsukuba, Japan
inoue@slis.tsukuba.ac.jp

Abstract - This paper presents a video advertisement system named "Jump-In-Ad." It captures a viewer's face in front of the system, and replaces the main character's face of the advertising videos with the viewer's face in real time. This advertising system is developed to attract people and to get their attention to the advertisement. From the experiment that compared the proposed system with a normal video advertisement system, it was demonstrated that more people looked at the "Jump-In-Ad" than the other.

Keywords: Digital signage, video advertisement, public display, interactive system.

1 INTRODUCTION

Public displays have become commonly used to replace conventional media such as posters and bulletin boards. These are called digital signage systems recently and are often used for advertisement. Advertisement by such display systems can be processed digitally unlike conventional paper advertisement. It can be a slide show, for example. However, many of these advertisements do not explore digital processing and simply plays a fixed clips repeatedly, which are not different from the conventional advertisement very much.

This paper presents an interactive advertisement system named "Jump-in-Ad" which embeds a viewer's face to the advertisement in real time. By embedding a viewer's face to an advertisement, his/her attention to the advertisement is expected to increase. The experiment to evaluate it is also presented.

2 RELATED RESEARCH

There are interactive advertisement systems. "Mirai Tube" displays information on the ceiling of a subway station when a person is detected at the concourse. The display changes according to the movement of the person. For example, it moves along with the walking [1]. GAS (Group-adaptive Advertisement System) displays advertisements that are appropriate to a group of viewers in front of the system by sensing the interpersonal distances of the group and estimating the group attribute [2]. In these systems, appropriate advertisements are selected and displayed according to the viewer. The proposed system "Jump-in-Ad," on the other hand, does not select an advertisement. Rather, it changes the advertisement in its content.

SIKUMI DESIGN [3] captures a viewer's face in real time and decorates the image with illustrations, but not the movie of the viewer's face. It is also used for advertisement but the effect is not known.

Figure 1: Appearance of the Jump-in-Ad system.

Capturing a user's figure is not a new idea. VideoPlace which captures a user's figure in a video was invented as an interactive media art in 1970's [4]. EffecTV is a toolkit to add visual effects to a movie in real time [5]. It is thought to be capable of applying to an advertisement, but not done. In DIM (Dive Into the Movie) [6], a 3D viewer's face is captured and used as the actor's face of a 3D movie. Although it is capable of capturing 3D face image, it uses 7 cameras to scan the image and cannot embed the image into the movie in real time.

There are applications of visual image or video of a user other than advertisement. Virtual Fashion adds decorations of a cloth, hair, and makeup to the movie of a user [7].

3 PROPOSAL OF "JUMP-IN-AD," AN INTERACTIVE ADVERTISEMENT SYSTEM

We propose an interactive video advertisement system named "Jump-in-Ad." It captures a viewer's face and embeds it to the advertisement in real time (Figure 1). It aims to attract the viewer's interest to the advertisement by applying the real time video processing technology to the advertisement itself. It is different from the previous viewer-adaptive systems that select an advertisement from the list of advertisements.

3.1 Hardware

The system consists of a 52 inch display panel for showing advertisements in public, a camera to capture a viewer's face, and a PC for processing the captured video and the

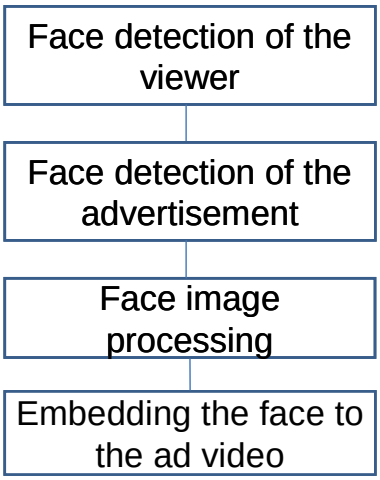


Figure 2: Software process of the system.

Figure 3: Mask image to trim the face shape.

advertisement. It needs a viewer as close as 2.5 m to detect the face.

3.2 Software

Figure 2 represents the software process of the system. The software is written in C++ and OpenCV on Windows.

First, the system tries to detect a face in the frames from the camera. It is done by every frame. Haar-like algorithm was used for the face detection. The number of faces, their locations in x and y, and their sizes in width and height are recognized. When multiple faces are detected, the biggest face is selected as the closest one to the system and used as the viewer's face currently.

The same face detection process is also executed in the advertisement. The size of the area to embed the face is recognized.

Next, the face image is extracted from the frame according to the location and the size data. The mask image to trim the face shape as ellipse is prepared. It is a white ellipsoidal figure with black background (Figure 3). The size of the mask is adjusted to the size of the area to embed the face.

Then the size of the viewer's face is adjusted to the size of the face area in the advertisement. The face image is trimmed by the mask image and is embedded to the area.

Figure 4 shows the original advertisement on the top and the processed advertisement that embeds the viewer's face at the bottom.

Example movies were prepared as the advertisement. They were 15 second short movies that show a face. This is

Figure 4: Top: original advertisement, Bottom: processed advertisement that embeds the viewer's face.

because of the copyright issue. The system does not require specially prepared advertisement.

4 EVALUATION

We have evaluated the system through a field experiment of comparing with a simple public display system.

4.1 Procedure

The proposed system and the simple public display system as the control system were used in the same condition. The control system displayed the same advertisement but was not interactive.

Two places were used in a university campus. One was an elevator hall and the other was a corridor (Figure 5). Each system was set an hour at a time. The systems were switched after an hour. The order of the systems was counterbalanced. The systems were set 8 times in 2 days all together.

The advertisement area was videotaped.

4.2 Result

Figure 5: Places of the experiment.

The measure of the effect of public advertisement is not yet standardized, which is thought to be different from the viewing rate of the television programs, for example. The number of people in front of the advertisement or the gaze at the advertisement has been used to measure the effect so far [8][9]. Taking this into account, the system was evaluated in the behavior of the viewers and the sojourn time in front of the system.

The viewers were categorized in the following three types according to their behaviors. They are A) Pass by without looking, B) Pass by with looking (face detected), and C) Stop and look (face detected).

Figure 6 shows the rate of the viewer's behavior in terms of these types. 3.7% was type C in the proposed system while it was 2.3% in the control system. More percentage of people stopped and looked in the proposed system. Also, 4.5% was type B in the proposed system while it was 3.5% in the control system. More percentage of people passed by with looking the advertisement in the proposed system.

The sojourn time in front of the system in type C and type B were shown in Figure 7 and Figure 8 respectively. It was 13.6 second in the proposed system and 13.8 second in the control system in type C, and 5.6 second in the proposed system and 4.1 second in the control system in type B.

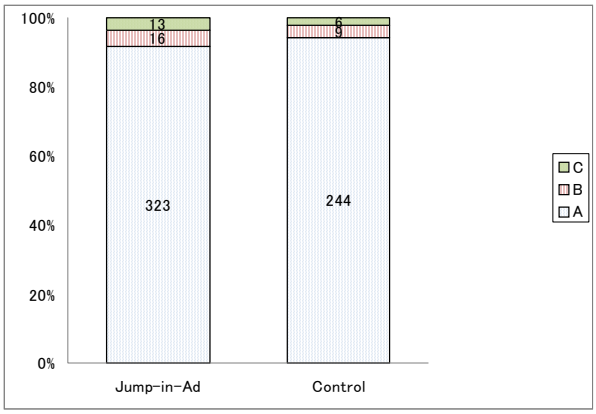


Figure 6: Viewers' behavior in the experiment.

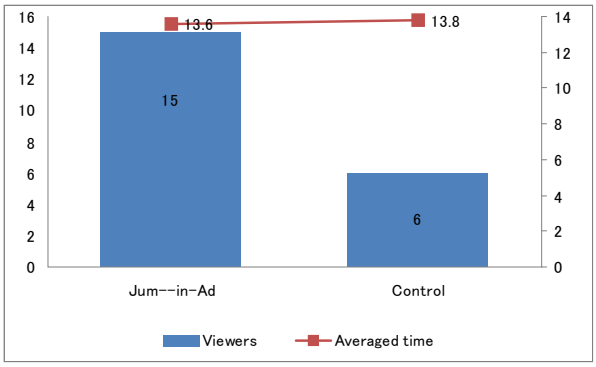


Figure 7: Sojourn time of type C people.

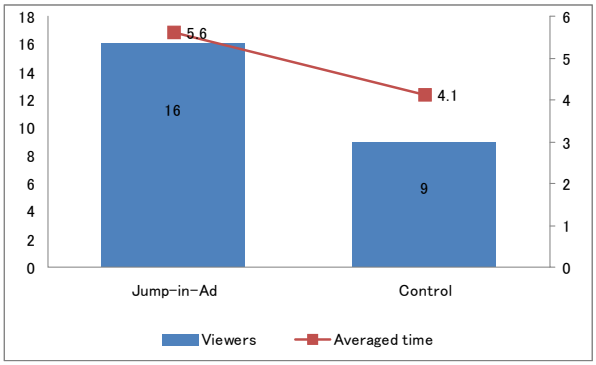


Figure 8: Sojourn time of type B people.

4.3 Discussion

The proposed system attracted more than the control system. Although the rate of type B and C was less than 10% of all the passers-by even in the proposed system, this rate cannot be discussed without the viewing ratio of this type of advertisement. The rate must also be dependent on the content of advertisement.

The proposed system did not work to type A people, who were not close enough to the system. The system works only when it detects the face of a viewer. A few people were found in the experiment who came close to the system when

they saw the system working to another viewer. The system must be effective when it is placed where there are some people around. To use the system effectively even where there are not many people, another method to let the distant passers-by notice the system may be desired.

5 CONCLUTION

We have proposed a video advertisement system “Jump-in-Ad” that embeds the viewer’s face into the advertisement in real time to attract more attention to the advertisement. Interactive advertisement can be adaptive to the viewer. While most previous systems select an appropriate advertisement from the list of advertisements according to the viewer’s profile, the proposed system changes the presentation of the advertisement with the viewer’s face.

From the field experiment, the proposed system could attract more interest than the simple display system. Future tasks were also addressed.

REFERENCES

[1] A. Shinohara, J. Tomita, and T. Kihara, Mirai Tube: A field trial of an interactive media in public space, IPSJ Technical Report, Vol.2006, No.14, pp.163-168 (2006).
[2] T. Inoue, and K. Heishi, GAS: a group-adaptive advertisement system for public spaces, IPSJ Journal, Vol.49, No.6, pp. 1962-1971 (2008).
[3] SIKUMI DESIGN, http://www.shikumi.co.jp
[4] M. Krueger, Artificial Reality, Addison-Wesley, MA, p.215 (1983).
[5] K. Fukuchi, S. Mertens, E. Tannenbaum, EffecTV: a real-time software video effect processor for entertainment, Entertainment Computing - ICEC 2004, LNCS, Vol.3166, pp.602-605 (2004).
[6] S. Morishima, "Dive into the Movie" Audience-Driven Immersive Experience in the Story, Trans. IEICE, Vol.91-D, No.6, pp.1594-1603 (2008).
[7] J. Hoshino, H. Saito, Building Virtual Fashion Simulator by Merging CG and Humans in Video Sequences, Trans. IPSJ, Vol.45, No.5, pp.1182-1193 (2001).
[8] C.J. Lee, et al., Attention Meter: A Vision-based Input Toolkit for Interaction Designers, Proc. CHI 2006 extended abstracts, pp.1007-1012 (2006).
[9] S. Minamitake, S. Takahashi, and J. Tanaka, A gaze information and movement trace acquisition toolkit to the digital advertisement, IPSJ Symposium Series, Vol.2010, No.4, SA12 (2010).

Science and mathematics education using robot simulations

Taiji Koike[1*], Takafumi Suzuki[1*], Keita Mori[1*], Akihiro Yorita[2*], Osamu Yuuki[4*]
Kouji Yoshida[3*], Naoyuki Kubota[2*], Tadanori Mizuno[4*], Kunihiro Yamada[1*]

*1 Tokai University, *2 Tokyo Metropolitan University
*3 Shonan Institute of Technology, *4 Shizuoka University

{9aejp005@mail.tokai-u.jp, yamadaku@tokai.ac.jp}

Abstract - We have developed four courses using special teaching materials, taking advantage of simulation experiments, and using game-like procedures with modern technologies to help students in primary and junior high schools understand science and mathematics and to augment their interest in these disciplines. These are based on commonly observed phenomena: (1) experiments help students feel and understand laws and principles of science and mathematics, and (2) amazing game-like operations often provide students with interest in these disciplines. We found that consecutive classes of these four courses in a short period promote their performance. For development, we have produced robots that simulate scientific experiments and perform game-like operations. Additionally, we have conducted experimental classes in primary and junior high schools to place our materials at the disposal of schools and to improve the materials by doing so. This research has been conducted under the auspices of Science Partnership Projects (SPP, a public research project) of the Japan Science and Technology Agency, an independent administrative institution.
Keywords: young students' aversion to science, robot, education, game

1. INTRODUCTION

We can recognize and appreciate recent developments in electronic devices, appliances, and vehicles that have been achieved through rapid development of microprocessors making full use of technologies of communication and control. Those technologies have transformed capabilities of mechanisms and electromechanical parts and have allowed them to mature over the years into higher systems. Consequently, boosted by growing human technology, great innovation has occurred in human–machine interfaces. In particular, remarkable innovations have been made in semiconductor and electronic parts. Hierarchical hardware and software mechanisms in conjunction with product development and computer aided design (CAD) systems have worked together effectively. Systems in product development are structured hierarchically and virtualized. Complicated products incorporating various technologies can be realized rapidly. That virtualization technology is indispensable in modern product and system development. Although that technology is a key in those endeavors, we only slightly recognize its exciting nature [1].

Students of the younger generation are moving away from exploration of science, perhaps because of virtualization technology. Although students of primary and junior high schools enjoy and resort to the convenience that products made using virtualization technology provide them, they might not be interested in overly complicated products themselves. Showing them the inside of a complicated product might be effective to help them become interested in science and related technologies, but we have not taken that road in this research. Instead, we have invented a course providing them simulation experiments using robots. The experiments, although not very spectacular, can be repeated at any time and can therefore offer students game-like fun. Consequently, we have sought to help them feel and know the laws and principles of science and mathematics. Our four courses are equipped with features of both simulation and game-like fun realized using robots.

2. PEDAGOGICAL CONSIDERATIONS

The trend of young people moving away from science has neither been defined clearly nor investigated sufficiently as a problem of elementary education systems in Japan. Although the importance of experiments in courses teaching science has been emphasized as a countermeasure against that trend, it has never been overlooked [2]. Not all problems in education are attributable to those of elementary education. We might think of three measures to settle any problem. The first is to remove the cause of the problem if found, the second is to alleviate problems irrespective of the knowledge of their causes. The third is to do nothing. We have taken the second option to activate science education.

The methodology of edutainment, with instruction accompanied by entertaining elements, was introduced long ago to enlighten and educate the public; in 1970, it was incorporated into radio programs. Such efforts are based on the idea that games enrich education [3, 4]. We can imagine education-giving robots of three types. The first is a robot contest, in which students compete with their own developed robots. In this type of activity, students learn many techniques and skills in developing robots that might work. They acquire, along the way, better capabilities of working together. Then, the second is a course using robots

where students learn the dynamics of robot behavior by watching them and learning further general knowledge and theory [5]. Finally in the third type, students learn with robots or robots teach them. Their robots become their friends and partners [6].

3. DEVELOPMENT OF TEACHING ROBOTS

We have developed four robots for courses in primary and junior high schools and tested them in experimental classes as activities of Science Partnership Projects (SPP) [7] of Japan Science and Technology Agency, an Independent Administrative Institution. Four robots are an Imagine car of the future robot (FUTURE VEHICLE), a SUMO ROBOT, a Parabola-throwing bio-pitcher (BIO PITCHER), and a MATH ROBOT. Features of these robots are presented in Table 1.

Table 1: Education programs using Robots

	FUTURE VEHICHLE	SUMO ROBOT	BIO PITCHER	MATH ROBOTE
Educational merit in science and mathematics	very small	small	medium	large
Teaching skills using game-like procedures	large	medium	small	very small

Table1 shows the ratio of the element.
large > medium > small > very small

In the table, the row of "educational merit in science and mathematics" expresses the educational performance of each robot in illustrating or teaching laws and principles of science and mathematics. The next row of "teaching skills by game-like procedures" expresses how each robot teaches skills involved in these disciplines in a game-like environment. These four robots contained the element of "educational merit in science and mathematics" and" teaching skills by game-like procedures" though it did not intend. "Bioloid Beginner Kit" [8] and "ROBOBUILDER" [9] were used for our experiments.

3.1 Future Vehicle

This is a robot simulation automobile that is very familiar to students. Safety devices embedded in automobiles are crucial for the eventual implementation of safe cars. The central technology in them is automatic control using microprocessors. We have developed a robot that can stop immediately before colliding with a wall or a human. Such devices for real automobiles are under development [10] and will soon be mounted in them. Our robot is hoped to help students understand the future automobile we are anticipating. Figure 1 and Figure 2 depicts the concept of our robot.

The robot is equipped with an obstacle-avoidance mechanism. Its sensor unit has an infrared transceiver module. Working together with these supportive devices, the PC maneuvers the robot through Zig-Bee communication. Students might simulate a driving situation, operating a PC by watching a picture transmitted from the camera mounted on the robot head. This driving simulation provides them with various effective edutainment, teaching safe driving techniques through two robots racing for example. This robot is intended to introduce sensor and control technology.

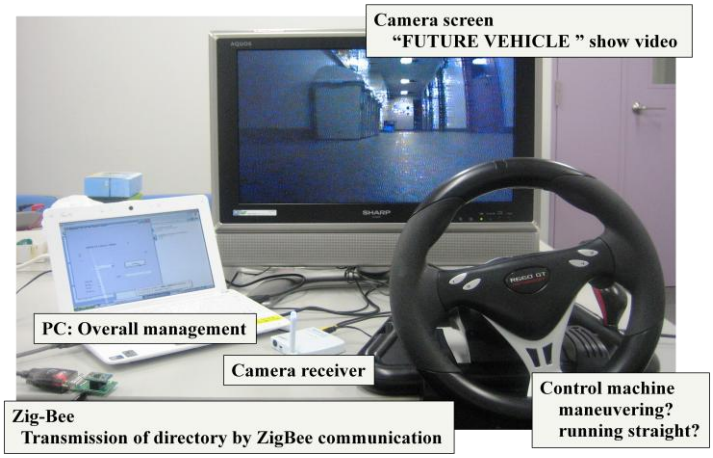


Figure 1: Configurations of future vehicles.

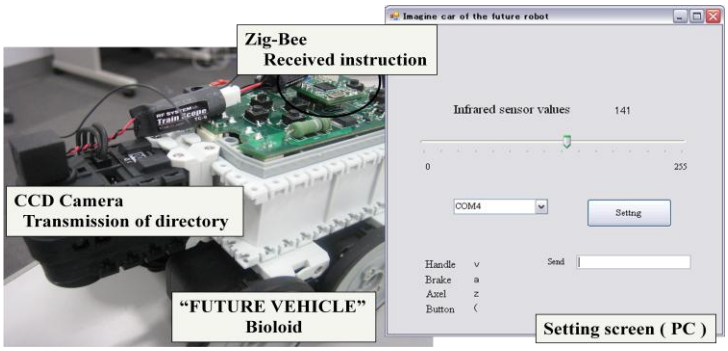


Figure 2: Configurations of future vehicles.

Figure 3: Air of the lecture [11]

The first goal of the course using the "FUTURE VEHICLE" is to help students learn the outline of infrared communication that is widely used in our electric appliances, such as the remote controller for a TV set. Students therefore understand that they are relying on invisible light rays. The second is to learn a method to develop automobiles that might forestall collision by themselves.

Students are asked to tune their robot vehicles equipped with obstacle avoidance mechanisms and a sensor unit that has not been adjusted properly yet. Third, students experience car racing and learn how to avoid accidents. Figure 3 shows the class scenery in the primary school.

3.2 Sumo Robot

Students often encounter robots, enjoying animated movies, for example. Such robots typically move exactly as human beings and their performance often outpaces that of human beings. However, practical robots perform only a few functions. Using this robot, we expect students to understand whether robots can move exactly as human beings. Students are asked to discuss that after manipulating the robot we prepared for them.

For this robot, we used "ROBOBUILDER", which walks with two legs and which can be maneuvered by students quite easily. "ROBOBUILDER" can memorize any movement of the body (motion), if one gives any motion to it using one's hand. This procedure is designated as "making a motion." Figure 4 shows the PC display when this is being performed.

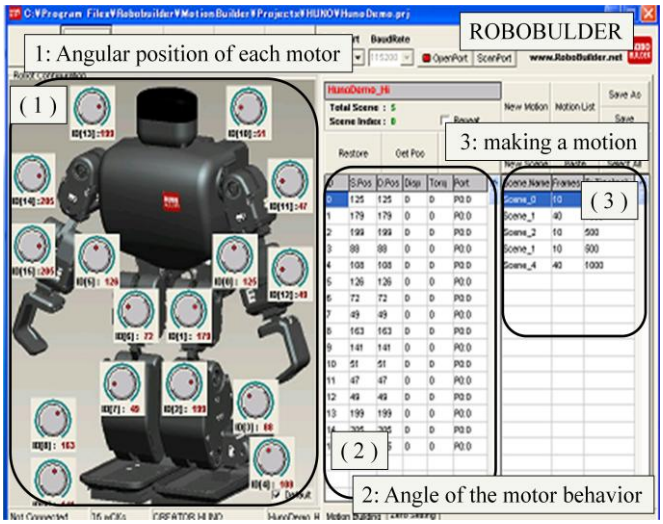


Figure 4: Sumo robot's programming motions on display.

The first goal of this Sumo robot is to make students recognize the difference of movements between those of robots and human beings. Students learn about the low visibility and poor maneuverability of robots and their difficulties in balancing. They then try to teach the robot some jujitsu moves (sumo wrestling). Meanwhile, they recognize the mobility and flexibility of human muscles. The second is to develop a robot itself making use of the "ROBOBUILDER". They program motions while considering the movement of the center of gravity and balance of the body. Finally they fight each other, maneuvering their robots using their own planned motions. Students might notice the difference in field of vision between the human eye and that of a robot because they fight only through wireless cameras mounted on robots. Figure 5 shows the lecture making scenery. Figure 6 shows the class scenery in the primary school.

Figure 5: Development scenery

Figure 6: Air of the lecture [12]

3.3 Bio Pitcher

By maneuvering this robot, students explore the skill of throwing by questioning how far they can throw. They learn about the high degrees of freedom of joint characteristics of humans and the difference of body movements between robots and human beings at the scene of throwing. We expect them to understand the function of force and the nature of parabolas through this experiment of throwing. Furthermore, we hope they realize that force is a vector, having two properties of magnitude and direction, and that the natural phenomena are controlled by the principles of science and mathematics. Figure 7 portrays a parabolic course of a ball thrown by the "BIO PITCHER".

Our bio-pitcher is equipped with a controller and actuator and can exert force on an object. Students recognize that force can deform things and change the condition of movement. Furthermore, as described above, they are to understand that force has two properties of magnitude and direction. The configuration of our BIO PITCHER is portrayed in Figure 8.

Students adjust parameters of throwing on their display and find, experimentally, the best setting to get the longest distance of throwing. Subsequently, they are asked to submit instinctive and logical estimations about throwing summarizing their group discussion. The adjustable parameters of throwing are the arm length, the actuator

angle, and the launch speed: 32 combinations of these parameters can be set. Figure 9 shows the class scenery in the junior high school.

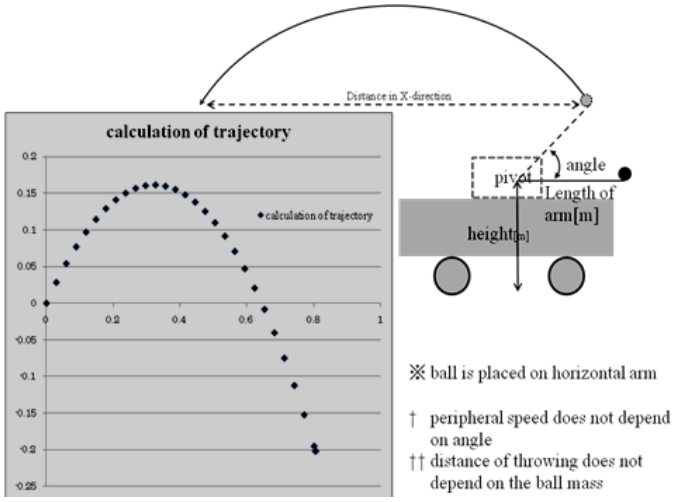


Figure 7: Parabolic path that a Bio-pitcher makes.

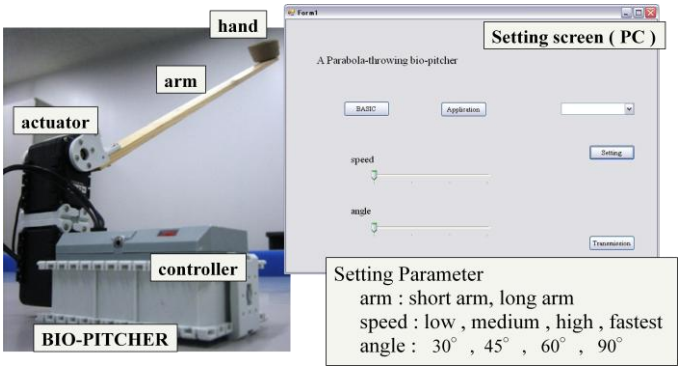


Figure 8: Configurations and setting parameter of Bio-pitcher

Figure 9: Air of the lecture

Students collect the throwing distance data for all these settings and discuss the differences between data and the two estimations presented above. It is necessary about the following two items if it is possible and learns. Simulation and reality differ, experiment is important.

3.4 Math Robot

Displacement, velocity, and time are some of the central concepts in physics and the relations among them described by mathematics. Using this robot, students learn the relations from the travel motion of the robot car. They feel and recognize the relations at two occasions. The first occurs when they are watching a running robot; the second is when they are preparing graphs demonstrating the robot car's travel motion. They might feel them as they like. It might be the proportional relation or the linearity of graph. If they expect the existence of a functional relation, then the experiments can be said to have been very productive.

The graph doesn't do the thing generally drawn in the line chart. Because a few error margins go out. However, the elementary school student uses the line chart because it doesn't learn it. Finding the relation becomes a purpose in the graph. Drawing style in graph is learning by the compulsory education accurately. Figure 10 Gap by graph.

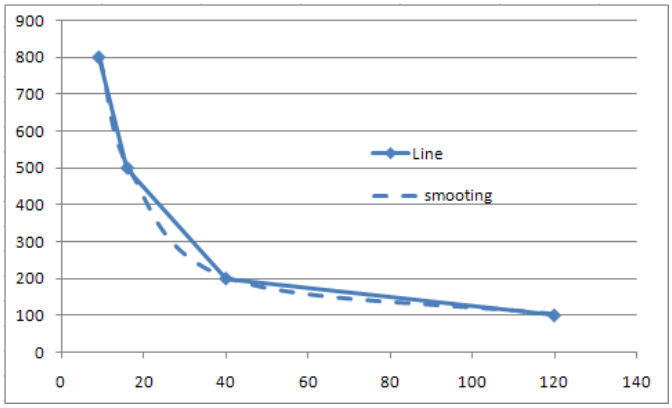


Figure 10: Gap by graph

Students are guessing, in daily life, the relation between displacement, velocity, and time by knowing the time interval required for walking a finite distance. Using the Math robot, they measure the displacement and time by themselves and present the data on graphs. They feel the relation among these physical quantities and then might reduce it to a simple mathematical expression. A set of these experiences is of the primary importance. In many classrooms, teachers are apt to adhere to teaching of the mathematical expression and will try to ask students to memorize it. This is, we believe, one of the causes of the trend among young students away from science. Our Math robot might make a linear run with uniform velocity and students might set a relation among displacement, time, and velocity on their PCs. Figure 11 shows our Math robot and the program controlling it. We merely ask students to show their data on a graph. At that time, we should not tell them about the type of graph and co-ordinate system. We should merely ask for some graphical representations. Perhaps they should be told that the common graph type used in these situations is only a rough representation. The instructor should not specify the type of graphical representation. If that were done, it would become another cause of young people's aversion to science. Figure 12 shows the lecture

making scenery.

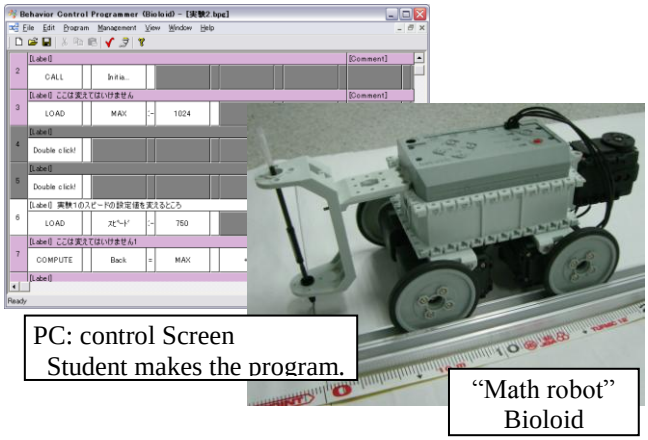


Figure 11: Math robot and program controlling it.

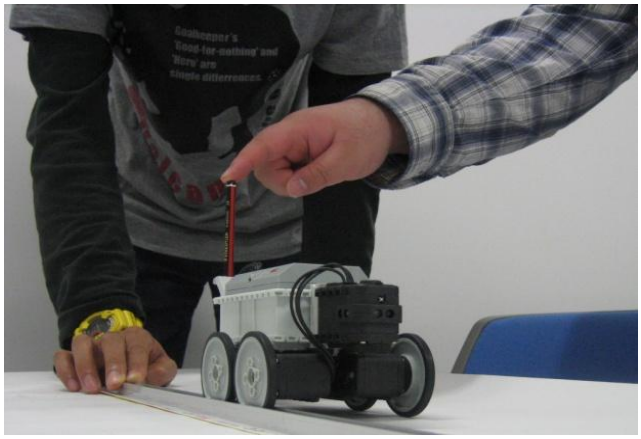


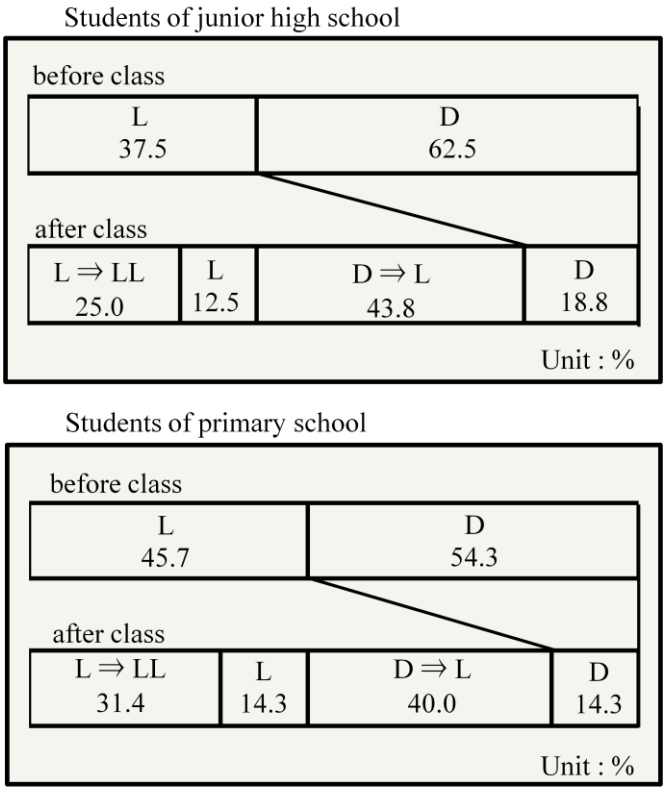
Figure 12: Development scenery

4. EXPERIMENTS AND EVALUATION

We gave experimental classes for 35 students of primary school and 16 students of junior high school and made a questionnaire survey of the change of interest toward science and mathematics after experiencing our experimental classes. The results of the survey are shown in Figure 13. Before the classes, the percentage of students in primary school who liked (L) science and mathematics was 45%. That of students in junior high school was 38%. The weighted mean of these is 43.1%: all are less than 50%, as expected. After the classes, the percentage of students in primary school who changed from "dislike" (D) to "like" was 40%, whereas that of students in junior high school was 43.8%. The weighted mean of these is 41.2%.

In summation, students voted the L element (D⇒L⇒LL) after the classes, they occupied 85.7% in primary school. That percentage in junior high school was 81.2%. The weighted mean of these two marks was 84.3%.

We must refine the art of asking questions next. Although we shall expect no great improvement by this first trial, we have not detected anything bad. Therefore, similar ones are worth further trial.



Like science and mathematics

L : like D : dislike

D ⇒ L : changed from "dislike" to "like"

L ⇒ LL : like science and mathematics

Figure 13: Students' interests changed after attending the course.

Some students in junior high school confessed that they attended our classes simply because their friends did so. Therefore, their attendance did not necessarily reflect a special curiosity about science and mathematics, and these data suggest the effectiveness of group learning. In a free description section of the questionnaire, students in primary school wrote that they found the role of science and mathematics in daily life and that they were interested in the mechanical and electronic parts used in robots. This report suggests polarization of the effectiveness of the robots: as an experiment or as a game. Some students said that they could accommodate the tough discipline of science and mathematics if they were helped by a robot. This might be a similar notion to that shown by students imitating their friends, and might reflect the recent loneliness of students in Japan. If so, it might be developing not only to an aversion to science and mathematics but also to an aversion to human beings, which we might also be concerned about. To our great surprise, many students of junior high school expressed skepticism and criticism of the idea of robots in the questionnaire before the class because of their imaginary nature. We are greatly interested in the underlying cause of that skepticism, whether it is the rough nature of robot systems or the surrealism of stories of robots losing touch with real life. However, students seemed happy touching

and using the robots. Robots might constitute an effective countermeasure against the trend of young people moving away from science and mathematics and might simultaneously provide them with fun.

The first key point is to abandon one-sided lectures and explanations and to ask students to repeat a simple measuring experiment. The second is to make them understand the meaning of that experiment, playing games packed with modern technologies. Do not teach theory and natural laws first. Summarize them simply after the experiment and playing game. Do not place importance on theories and natural laws. Never ask students to memorize them. That surely enhances the aversion of young people to science and mathematics.

It is important as a countermeasure against aversion to science and mathematics to lead students to infer something from simple repeated actions. That stance is valuable by itself as well and encourages students to pursue their definite goal. To guide students into that direction is the very role of education. We have felt some success in this context. After experiments, students exercised great enthusiasm in calculation, discussion, and presentation. Students of primary school have been very active and excited, and have shown strong concentration. Students of junior high school have shown excellent persistence in their logical progression to raising points at a question and answer session given after the presentation.

5. SUMMARY

We have developed four teaching materials to induce students in elementary education to be more interested in science and mathematics. Each covered the elements of experiments and games in different weight. Then we had experimental classes making use of those teaching materials in primary and junior high schools. We made questionnaire studies of the change of interest toward science and mathematics before and after our experimental classes, finding a positive trend. Providing four courses consecutively in a class yielded unexpected success. Students operated different robots, made group discussions and attended presentations. They were filled with tense excitement and made a unified effort with group members. They struggled and fought with science and mathematics, and then became winners.

In this article, two problems were left. First, It questioned immediately after and immediately before. It was very a great result. Next stage, we want to investigate and to examine whether the interest and the concern continue. Second, we have developed four teaching materials to induce students in elementary education to be more interested in science and mathematics. But they were only continuously executed. Next stage, how do the concern and the concern change while teaching each element (“Technical element” and “Educational element ") from a strong lecture to a weak lecture? This will become one of the research topics.

ACKNOWLEDGEMENT

We appreciate Principal Tomoo Kume, Nobuaki Nakano, and students of Mikkabi Hamamatsu Municipal Primary School, Shizuoka Pref., all of whom accepted our experimental classes under the auspices of Science Partnership Projects (December 3–4, 2009). Furthermore, our appreciation is extended to Dr.Ryo Ishii and his students of Junior High School Division of Takanawadai High School affiliated with Tokai University (December 19, 2009). Komiyama, SPP Head Office, Japan Science and Technology Agency, an Independent Administrative Institution, has long taken care of office procedures, which we appreciate deeply.

REFERENCES

[1] K.Yamada, K.Yoshida, M.Kojima, T.Matsumura, T. Mizuno (2009) New System Structuring Method That Adapts to Technological Progress of Semiconductors, KES 2009, Part II, LNAI5712, pp.773-782.

[2] K.Kamitori et al. (2004) Report on Reviewing Experiments in Science Class at Public Junior High School in Three Prefectures of Kinki-district, Annual Report of Science Education, Osaka Kyoiku University, No. 25, p.21-32.

[3] K.H.Veltman, Edutainment, Technotainment and Culture, Veltman, K. H., Città Annual Report, 2003.

[4] M. Ito (2004) Engineering plays: Children’s software and the cultural politics of edutainment, Discourse 27 (2), pp.139-160.

[5] N. Kubota, Y. Wagatsuma, S. Ozawa (2009) Intelligent Technologies for Edutainment using Multiple Robots, Proc. (CD-ROM) 5th Int’l. Symp. Autonomous Minirobots for Research and Edutainment (AMiRE 2009), 195-203.

[6] A.Yorita, T.Hashimoto, H.Kobayashi, N.Kubota, (2009) Remote Education based on Robot Edutainment, Proc. (CD-ROM) 5th Int’l. Symp. Autonomous Minirobots for Research and Edutainment (AMiRE 2009), 204-213.

[7] http://spp.jst.go.jp/

[8] http://www.robotis.com/zbxe/main

[9] http://roboone.shop12.makeshop.jp/html/newpage.html?code=13

[10] Safety Technology of Automobile, Web page of Nissan Motors, http://e2a.jp/trend/070627.shtml

[11] Chunichi Newspaper,title is “The science and arithmetic are interesting.” 2009.12.05

[12] shizuoka Newspaper, title is” The operation of the robot is freedom” 2009.12.05

Session 3: Network1
(Chairs Yoshia Saito and Tomoya Kitani)

Technology for Multi-database Virtualization in a Ubiquitous Computing Environment

Yuta Watanabe[*], Keisuke Syoubu, Hiroshi Miida, Yuji Wada[**],
Jun Sawamoto[***], and Takashi Katoh[***]

[*]Graduate School of Information Environment, Tokyo Denki University, Japan
[**] Department of Information Environment, Tokyo Denki University, Japan
[***] Faculty of Software and Information Science, Iwate Prefectural University, Japan
[*]DEN03992@nifty.com, [**] yujiwada@sie.dendai.ac.jp, [***] sawamoto@iwate-pu.ac.jp

Abstract -Currently, large amounts of data are collected from ubiquitous sensor network environments. By using data mining techniques, the discovery and analysis of the knowledge and trends hidden in this data have become important in many applications, such as business decision-making. This data resides in a wide variety of distributed databases (i.e., a multi-database). A problem with the multi-database, however, is that despite an inherent aspiration to concentrate on analysis and rule extraction, data mining analysts nevertheless spend large amounts of time on such tasks as database selection and data collection to prepare for the data mining process. Therefore, in this study, to reduce the burden on data analysts, we developed database virtualization technology that makes it possible to use a multi-database in a ubiquitous computing environment as if it were a single database.

Keywords: Internet, virtual environment, web service, ubiquitous information processing, database, XML

1 INTRODUCTION

Nowadays, large amounts of data are being collected from ubiquitous sensor network environments, and the knowledge and trends hidden in this data, which can be discovered and analyzed through the use of data mining techniques, are becoming important in areas such as business decision making. This data resides in multiple types of distributed databases (heterogeneous databases). A problem with heterogeneous databases, however, is that despite the inherent desire to concentrate on analysis and rule extraction tasks, data mining analysts end up spending large amounts of time on such tasks as database selection and data collection in preparation for the data mining process.

Therefore, in order to reduce the burden on the data analysts, this study aims to develop database virtualization technique that enables heterogeneous databases to be used in a ubiquitous computing environment as if they were a single database.

This research was supported by a Grant-in-Aid for Scientific Research C (Subject No. 20500095).

Due to differences in data models and vendors, various types of heterogeneous databases exist. In databases employing different data models, data is represented differently, e.g., in table or XML format, and both the data storage method and query language are also different. Moreover, even among databases based on the same model, since the products will have been developed by different vendors, differences in functions and SQL will exist even for the same relational database (RDB). To simultaneously handle databases with different models or vendor functions, an application program interface (API) is required for each database (DB), and in the case where system specifications are changed during operation, the greater the impact of that change and the more work will be required. Virtualization of such databases and standardization of their procedures enable the burden on analysts and the development costs to be reduced, and it is expected to increase operating efficiency. Here, we describe an API for a virtualization DB.

2 RELATED RESEARCH

Several studies relating to database virtualization research have previously been published. In reference [1], using a group of heterogeneous databases connected to a wide-area network as an information source, a method is proposed for implementing a system to actively deliver information to users in a mobile computing environment. Actual data mapped from local database groups is combined to construct a heterogeneous meta-database, allowing search functions to be implemented. However, since the objective is data acquisition, there is no update function. Reference [2] similarly describes the virtualization of heterogeneous data, whereby data from various sources such as RDBs, Web services, ERP and CRM business applications can be accessed and integrated in real-time via a data virtualization system. However, only a virtualization function for RDBs from different vendors is supported, and there is no support for virtualization of DBs based on heterogeneous data models. Reference [3] similarly describes a module known as a wrapper that allows accessing and integrating data from various sources such as RDBs, the Web, and Excel files.

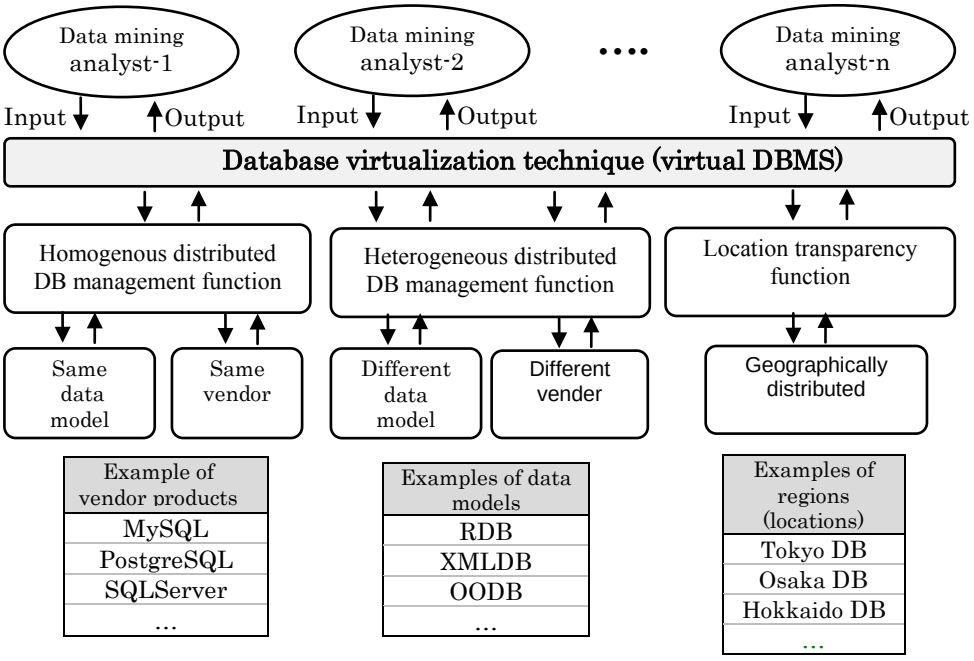
In this study, based on a common XML schema described in references [4, 5, 6], a virtual database is developed in

which various heterogeneous databases can be accessed and manipulated in response to a single query.

3 TECHNIQUE FOR DATABASE VIRTUALIZATION

Figure 1 shows the layers of the database virtualization technique described in references [4]. So that the user is not made aware of differences in structure or location, homogenous and heterogeneous distributed database management functions and a location transparency function are incorporated to provide a single database management system, capable of flexibly supporting any type of database, as described in references [5].

Figure 2 shows an overview of the database virtualization technique. In references [6], to realize virtualization of DBs based on such heterogeneous data models as XML database (XMLDB) and object oriented database (OODB) [7], an expressive and portable XML schema was utilized. This is a first step, and if required, UML, class diagrams and the like may also be used. Schema information for the different DBs is expressed as a single XML schema, based on which the user issues a single query and the virtual DBMS accesses each DB according to that query. A single result is returned to the user.



MySQL: Sun Microsystems
PostgreSQL: PostgreSQL Global Development Group
SQLServer: Microsoft

Figure 1: Technique for database virtualization.

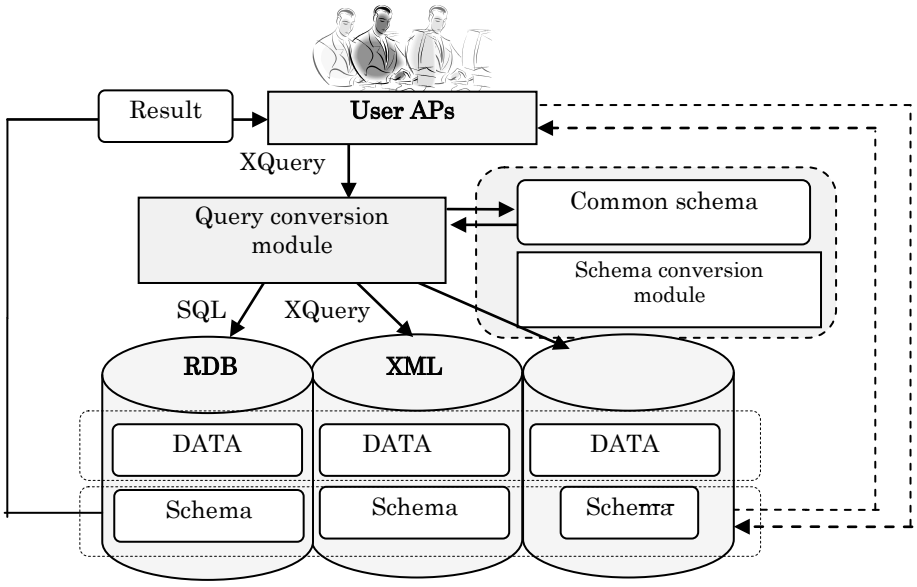


Figure 2: . Overview of database virtualization method.

4 VIRTUAL DATABASE SCHEMA DEFINITION FUNCTION

4.1 Common Schema Structure

A common schema provides the user with information about the structure of the virtual database, and is also used for checking query statements and constraints. As an example, Figure 3 shows the structure expressed by a common schema for a RDB and XMLDB.

The root element is located at the top, and the level below contains elements indicating various types of databases, for example, DB1 and DB2. The elements of the next lower level contain, in the case of a RDB, the database name, and in the case of a XMLDB, the collection path. Then, stored in the next lower level, in the case of a RDB, is an element indicating the table structure, and in the case of a XMLDB, the document structure. The RDB table structure first has an element indicating the table name, followed by the "row" elements listed below. The next lower level contains an element indicating the column name.

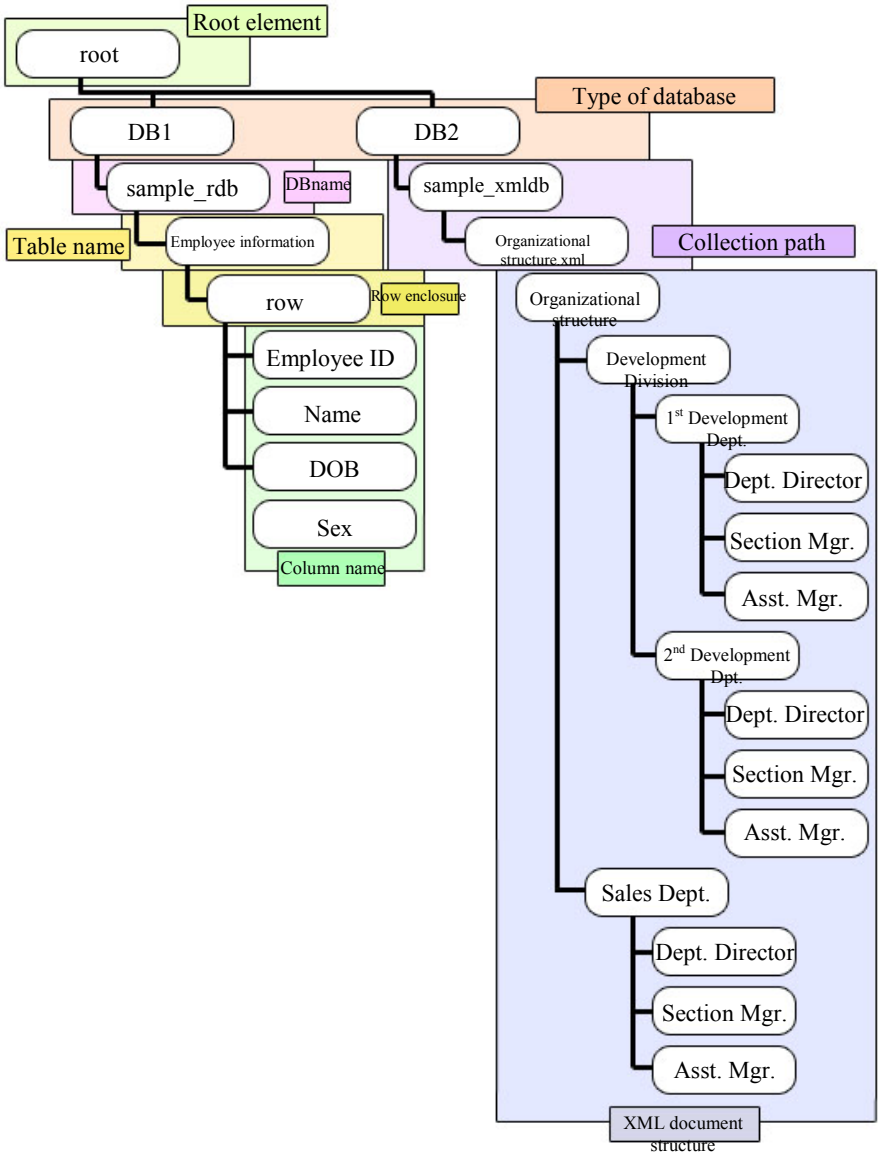


Figure 3: Example of structure of common schema.

Figure 4 shows the structure, expressed in XML, of DB1 in Figure 3. The table name "Employee information" is used as the root element, and below, data is entered into the columns of "Employee ID", "Name", "DOB" (Date of Birth) and "Sex" enclosed by "row" elements. The locations enclosed within this "row" form one row of the table. Similarly, data enclosed by "row" elements of the second and subsequent rows are arranged as sub-elements of the element indicating the table name. Constraint information is provided using <xsd:annotation> and <xsd:appInfo> as information-oriented comments. The main details regarding the outputting of constraints are as follows. (Primary key (unique) constraint: < r:index index-key="column name" primary(unique)="yes">, external key constraint: <f:key

fkey-column=”column name” ref-column=”reference column” ref-table=”reference table” rule-delete="operation at delete” rule-update="operation at update" />,check constraint: <r:check check-column=”column name” rule=”rule”>)

```
<Employee information>
        <row>
  <Employee ID>1</Employee ID>
    <Name>Taro Yamada</Name>
    <DOB>1970-03-04</DOB>
    <Sex>Male</Sex>
  </row>
  <row>
  <Employee ID>2</ Employee ID>
    <Name>Shinji Kaneko</Name>
    <DOB>1970-03-04</DOB>
    <Sex>Male</Sex>
  </row>
  …
</Employee ID>
```

Figure 4: Example of XML schema.

4.2 Common Schema Creation

Here, as a sample schema used in this operation example, a database containing the following information is created and an XML schema is then constructed using PostgreSQL. Similarly, to verify the basic constraints, in a column, the primary key constraint is assigned to “Employee ID, Affiliation ID”, the unique constraint is assigned to “Affiliation”, the external key constraint is assigned to “Affiliation ID”, the check constraint is assigned to “salary”, and the “NotNull” constraint is assigned to the “name”.

- CREATEDB EmployeeDB;

- CREATE TABLE Employee Table (
  Employee ID int PRIMARY KEY,
  Name varchar(50) NOT NULL,
  Salary int CHECK(0 < salary),
  Affiliation ID int REFERENCES Affiliation table (Affiliation ID)
  ON UPDATE CASCADE ON DELETE CASCADE);

- CREATE TABLE Affiliation table (
  Affiliation ID int Primary Key,
  Affiliation varchar(5) UNIQUE);

Figure 5 shows the XML schema that is output based on the above RDB schema. It can be seen that the column name and constraint information are outputted.

5 VIRTUAL DATABASE QUERY FUNCTION

A query inputted by a user is a single query to a virtual database rather than an individual query to each database. Possible candidates for user-issued queries are SQL, XQuery[8], and virtual database original queries. In this study, the common schema expressing the entire virtual database structure is described in XML format, and since this is presented to the user as the structure of the entire virtual database, XQuery was chosen as the type of query to be issued by a user to the virtual database. This is a first step, and in the future, when OODBs as well as RDBs and XMLDBs are supported, if the required query cannot be expressed with XQuery or is complicated and difficult, defining an original query for a virtual database may also be considered.

```
<!—Omitted-->
      <xs:element name=" Affiliation table"
type="Affiliation table Type"/>
      <xs:element name=" employee table "
type="Employee table Type"/>
   </xs:sequence>
</xs:complexType>
<xs:complexType name="Affiliation table Type">
   <xs:annotation>
      <xs:appinfo>
         <r:index index-key=" Affiliation id"
primary="yes"/>
         <r:index index-key=" Affiliation "
unique="yes"/>
      </xs:appinfo>
<!— Omitted -->
      <xs:complexType>
         <xs:sequence>
            <xs:element minOccurs="1" name="
Affiliation id" r:nullable="false"
            <xs:element minOccurs="0" name="
Affiliation " r:sqltype="varchar"
<!—Omitted-->
<xs:complexType name=" Employee table Type">
   <xs:annotation>
      <xs:appinfo>
         <r:index index-key="Employee id"
primary="yes"/>
  <r:check check-column="Salary" rule="0 &lt;
&quot;salary&quot;"
  <r:fkey fkey-column=" Affiliation id" ref-column="
Affiliation id" ref-table="
```

Figure 5: Example of common schema.

5.1 Query Conversion

When both a RDB and XMLDB are used at the same time, differences will exist in the query language used for each database, e.g., SQL for the RDB and XQuery for the XMLDB. Moreover, differences will exist in the data model since the RDB data is managed in a table-like structure, and the XMLDB is managed in a tree structure. Consider, for example, the case in which data as in Figure 6 is stored in the RDB and data as in Figure 7 is stored in the XMLDB.

Then, consider the case in which, based on these two data sources, it is desired to learn about the Dept. Director of the 1st Development Dept. In this case, a query such as the following will be issued.

Table name: Employee information

Employee ID	Name	DOB	Sex
1	Taro Yamada	1970-3-4	M
2	Shinji Kaneko	1975-4-2	M
3	Yasuko Oota	1974-1-9	F
4	Reika Yasuda	1980-6-25	F
5	Isamu Kirishima	1948-10-16	M
6	Koki Oonishi	1958-8-2	M
7	Kazuo Yamaguchi	1970-3-4	M
8	Kaoru Saitoh	1983-2-12	F
9	Yasushi Yamagishi	1986-3-5	F
10	Kenji Kimura	1986-6-2	M
11	Hiroki Yuuki	1986-9-30	M

Figure 6: Example of employee information table.

```
        Organizational Structure.xml
< Organizational Structure >
  <Development Div.>
    <1st Development Dept.>
      <Dept. Director>3</ Dept. Director >
      <Section Mgr.>2</ Section Mgr.>
      <Asst. Mgr.>7</ Asst. Mgr.>
    </1st Development Dept.>
    <2nd Development Dept.>
      < Dept. Director >5</ Dept. Director >
      < Section Mgr.>6</ Section Mgr.>
      < Asst. Mgr.>10</ Asst. Mgr.>
    </2nd Development Dept.>
    <Sales>
      < Dept. Director >3</ Dept. Director >
      < Section Mgr.>4</ Section Mgr.>
      < Asst. Mgr.>8</ Asst. Mgr.>
    </Sales>
  </Development Div.>
</Organizational Structure>
```

Figure 7: Example of XMLDB data.

● To learn the name of the Dept. Director of the 1st Development Dept.

・Obtain employee ID and name list from RDB

```
SELECT  Employee ID, Name
                    FROM EmployeeInformation;
```

・Obtain targeted employee ID from XMLDB

```
let $a := document("Organizational Structure.xml")
        /Organizational Structure/Development Dept./1st
Development Dept./Dept. Director
    return $a
```

The execution results from this query are as shown in Figure 8 and Figure 9.

○RDB

Employee ID	Name
1	Taro Yamada
2	Shinji Kaneko
3	Yasuko Oota
4	Reika Yasuda
5	Isamu Kirishima
6	Koki Oonishi
7	Kazuo Yamaguchi
8	Kaoru Saitoh
9	Yasushi Yamagishi
10	Kenji Kimura
11	Hiroki Yuuki

Figure 8: Example of employee ID and name list.

○XMLDB

```
<Dept. Director>3</Dept. Director>
```

Figure 9: Employee ID of Dept. Director of 1st Development Dept.

Results similar to the above are output in response to each query; however, the results that the user ultimately desires cannot be output based on the results from the RDB and XMLDB, and the user must personally compare those results. Moreover, because each query language is different, the user must be aware that multiple queries (SQL and XQuery in this case) are issued according to the DB used. Additionally, the user must also ascertain which data is stored in which database. These requirements place a large burden on the user, and as a result, certain tasks become very time consuming.

When implementing this example with a virtual database, the user will issue the following type of query.

● To learn name of Dept. Director of 1st Development Dept.

```
for $employee in common-schema()/RDB
      /sample_rdb/Employee information/row
   for $manager in common-schema()/XMLDB
      /sample_xmldb/Organizational structure.xml
       where $employee/Employee ID
      = $manager/Organizational structure/Development
Div./1st Development Dept./Dept. Director
   return $employee/name
```

The execution results obtained in response to this query are as shown in Figure 10.

```
<name>Yasuko Oota</name>
```

Figure 10: Execution results.

When the actual data was checked, the content of the /Organizational structure/Development Div./1st Development Dept./Dept. Director element from the XMLDB data was found to be "3." Then, by looking up the name of the individual having an employee ID of "3" in the data of the "Employee information table" in the RDB, the name "Yasuko Oota" can be verified. Thus, it is found that the required information was obtained correctly in response to the query.

5.2 Common Schema Query

The common-schema() statement in the query is a newly defined function for the virtual database, and is used to indicate the root of the common schema. By specifying, after the common-schema() statement, the path for obtaining the required values, the user is able to obtain RDB data or XMLDB data depending on the format that the data is stored in. These values are converted to object format, and set as the QueryArg object (as shown in Figure 11), which represents a variable. In this example, RDB data is set as the object representing $employee and XMLDB data is set as the object representing $manager. Further, by specifying the path and using a "where" clause to specify a condition for the value stored in this variable, RDB and RDB, RDB and XMLDB, and XMLDB and XMLDB values can be compared, and only that content which satisfies the condition can be returned. When attempting to obtain this result, the user sees only a structure indicating a common schema, and is not aware of the actual databases existing behind the virtual database. Of course, the individual queries, such as SQL and XQuery, for the actual databases are not generated. Thus, the values acquired from the actual DBs can be compared with one another, without the user being aware of their differences.

In this manner, the use of a virtual database enables RDB and XMLDB values to be acquired and utilized without the user being aware of the difference among the databases.

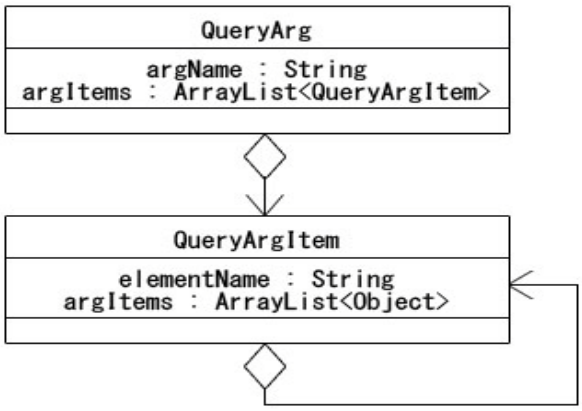


Figure 11: Variable objects in a query shown

6 VERIFICATION PREPARATION FOR DATA MINING APPLICATION FUNCTION

6.1 Verification from User Interface

The effectiveness of a virtual database that uses data mining is verified. Here, the tool described in reference [9] is treated as a data mining support tool, which enables the results of the mined data to be visualized, and helps analysts obtain a true understanding of the situation.

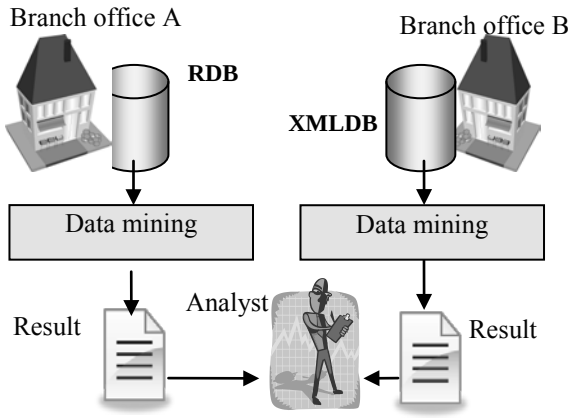


Figure 12: Analysis from multiple sources.

First, for the purpose of verification, transaction data containing the purchase history of a convenience store is used as the data to be analyzed. Consider the case in which the data is stored in a RDB at branch office A and in a XMLDB at branch office B, as shown in Figure 12. Usually, when performing the analysis, data is output from both databases and the outcome results must be compared using a data mining tool.

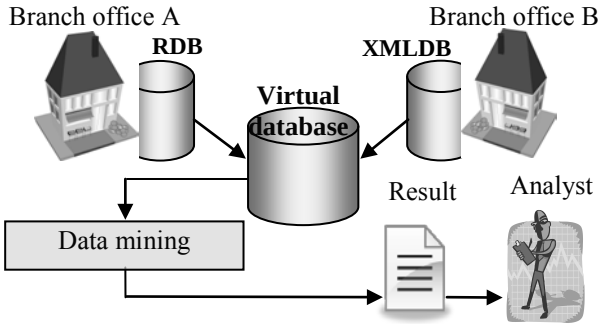


Figure 13: Analysis when using a virtual Database.

In a similar scenario, Figure 13 shows the way in which data mining is performed using a virtual database. As can been seen in the figure, since data is output from a virtual database formed from multiple databases, there is only one data output, rather than multiple outputs. Thus, a single result can be extracted by mining this data. In other words, there is no need to perform multiple data mining operations, and there is no need to compare the results. For example, even when using the tool described in reference [9], since the result can be output as a single comprehensive result

rather than as multiple results, the outcome can be discovered easily by analyzing and visualizing data from the two branch offices.

6.2 Verification from Application Interface

The API for a virtual database is described below. In this study, a virtual database using XQuery was developed as a first step. However, the effectiveness of several other patterns should be compared and reviewed. The proposals are as follows.

- **Proposal 1. XQuery + extended functionality**

This proposal is to support XQuery plus extended functions not provided by XQuery, as shown in Figure 14. Although XQuery allows flexible queries, there is a problem in that the statement length will increase in the case of a complex query. Conceivable extensions include a control or update function for constraints associated with the RDB.

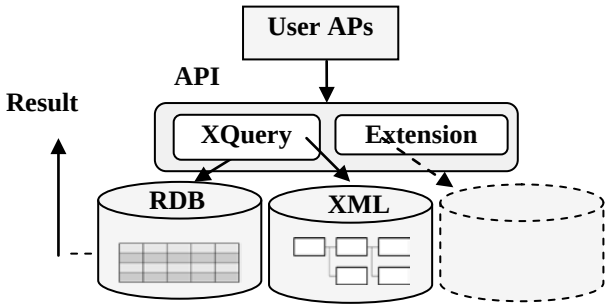


Figure 14: API (XQuery + extension).

- **Proposal 2. SQL + extended functionality**

This proposal is to support SQL plus functions not provided by SQL, as shown in Figure 15. SQL can be written simply with short statements, but it lacks flexibility. Therefore, as extensions, the addition of statements such as a "For" statement, a function for issuing queries to such elements as a XMLDB, and a function for discriminating between elements having the same name are considered essential.

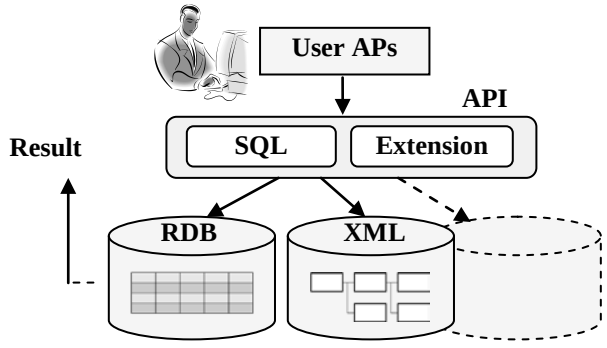


Figure15: API (SQL + extension).

- **Proposal 3. New query language**

The two types of queries based on the above proposals have advantages and disadvantages. Therefore, as shown in Figure 16, this proposal is to develop a completely new query language, rather than use an existing language. It is thought that by considering various patterns, an appropriate query can be developed.

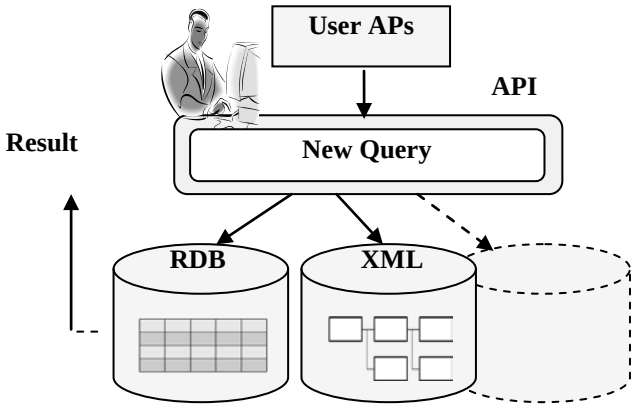


Figure 16: API (new query language).

7 CONCLUSION

Conventionally, when acquiring data from multiple types of DBs, a user has to issue a query corresponding to the DBMS, and then use the results from that query to infer the desired information. The newly developed virtual database enables the user to acquire and compare information as if only a single database were being accessed, without the user being aware of the differences among the databases. As part of this development work, an API for a virtual database was designed. In the future, we plan to compare and review this API, to consider which patterns are most suitable, and then implement and verify those API patterns.

ACKNOWLEDGEMENT

This research was supported by a Grant-in-Aid for Scientific Research C (Subject No. 20500095).

REFERENCES

[1] K. Mori, S. Kurabayashi, N. Ishibashi, and Y. Shimizu, Method of Sending Information Actively Reducing User Information Load Dynamically in a Mobile Computing Environment, DEWS2004 (March 2004).
[2] teiid, http://www.jboss.org/teiid, Red Hat.
[3]DB2:Information Integrator V8.1, http://www.jpgrid.org/documents/pdf/WORK4/sugawara_ws4.pdf
[4]Y. Wada, Y. Watanabe, J. Sawamoto, and T. Katoh, Database Virtualization Technology in Ubiquitous Computing, Proc. 6th Innovations in Information Technology (Innovations'09), pp.170-174 (2009-12).
[5] Y. Wada, Y. Watanabe, K. Syoubu, J. Sawamoto, and T. Katoh, Virtualization Technology for Ubiquitous

Databases, Proc. 4th Workshop on Engineering Complex Distributed Systems (ECDS 2010), pp.555-560 (2010-02).

[6] Y. Wada, Y. Watanabe, K. Syoubu, J. Sawamoto, and T. Katoh, Virtual Database Technology for Distributed Database, Proc. IEEE 24th International Conference on Advanced Information Networking and Applications Workshops(FINA2010), pp.214-219(2010-4).

[7] OfficeM, Encyclopedia SQL99, SYUUWA System, pp.462 (2002).

[8] S. Buxton, J. Milton, K. Shibano, K. Yamahira, T. Kodera. and M. Tsuchida, XQuery Querying XML, XQuery, XPath,, and SQL/XML in Context (Programmer's SELECTION), SHOEISYA, pp.669 (2008).

[9] H. Miida, and Y. Wada, Development of Comparison and Analysis Supporting Tool of Data Mining, FIT 2009, D-030, pp. 195-198(2009).

Graph Construction Method and Restoration Method for Application Level Multicast Using Network Coding

Kenji Otake*, Yoh Shiraishi** , and Osamu Takahashi**

*Graduate School of Systems Information Science, Future University Hakodate, Japan
**Future University Hakodate, Japan
{g2109008, siraisi, osamu}@fun.ac.jp

Abstract –This paper proposes a graph construction method and restoration method for application level multicast (ALM) using network coding (NC). We tested and evaluated the proposed methods.

Multicast communications have received a lot of attention in recent years because network load is increasing. There are two types of multicast communications: IP multicast and ALM. More flexible ALM has attention focused on it and is an active area of research. Network efficiency of ALM is lower than that of IP multicast. Therefore, we focus on NC to improve communication efficiency of ALM.

When NC is applied to ALM, a special path design is needed unlike usual ALM. In addition, nodes often leave ALM because it consists of end hosts. We developed a graph construction method and restoration method for when NC is applied to ALM. We conducted an experiment to evaluate these methods. We evaluated delay and throughput to prove that our proposal is better than existing ALM. In addition, we measured overhead when paths were restored when nodes left and verified the method's practicality.

Keywords: Multicast, P2P, Overlay network, Network coding

1 BACKGROUND

Multimedia traffic, such as video and audio, has been increasing in recent years because hosts that access networks are also increasing. However, these contents are often spread redundantly. Thus, multicast communication is currently receiving a lot of attention. There are two type main types of multicast communications: IP multicast and ALM [1]. IP multicast provides multicast communication on a network layer and has high communication efficiency. However, all routers need to accept multicast, so this has low applicability. Therefore, ALM, which provides multicast communication on an application layer, is receiving a lot of attention and becoming an active area of research. Unlike IP multicast, ALM can produce multicast communication without network support for multicast. However, this causes local link stress in ALM because ALM uses unicast communication on the network layer. Therefore, we focus on NC [2].

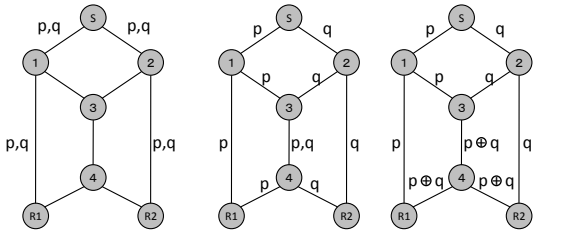
NC encodes packets in a network. It effectively decreases network load and improves the network resources. When NC is applied to ALM, a special path design is needed. Therefore, we developed a graph construction method. In addition, we developed a graph restoration method for solving the frequently occurring problem of nodes leaving. We conducted an experiment involving each proposal and evaluated performance and the method's practicality.

2 RELATED WORK

2.1 Network Coding

NC encodes packets in a network and effectively decreases network load in multicast communication.

Consider the network in Figure 1, where source S wants to deliver the stream of messages p and q to both R1 and R2. If it uses usual multicast communication, the 2 packets go through all used links (Figure 1(a)). Two packets can go through different links to disperse network load (Figure 1(b)). However, in this case the link between nodes 3 and 4 becomes a bottleneck because 2 packets, p and q, go through this link. As a result, network load increases locally and throughput does not progress. Node 3 encode (for example XOR) packets arrived at this node, make a new packet, p+q, and transfer it with NC. Then the bottleneck is removed because one packet goes through the link between 3 and 4 (Figure 1(c)). This packet, p+q, contains information about both p and q. Thus if a node gets this packet and a p or q packet, this node can decode these packets and extract the other packet. Therefore, reception nodes R1 and R2 extract p ore q from p+q and source S could transfer p and q. Network load is dispersed, and used band decreases with NC. Routers need to accept the encoding packet to provide this. The router's process in the current network is limited under the network layer. Therefore the new router accepting NC must spread to provide NC in the current network.



(a) In case of normal multicast (b) In case trying to distribute load (b) In case using network coding

Figure 1: Multicast communication and NC

2.2 Application Level Multicast

ALM is a multicast communication on an overlay network built by the end host's P2P connection on an IP network. Unlike IP multicast, end hosts transfer and copy packets in ALM, so a network does not require special acceptance. However, ALM provides multicast communication on an overlay network with using unicast on the network layer. Thus, multi-passes are built on a single physical link. Also, ALM is less efficient than IP multicast. In addition, end hosts are generally less reliable than network equipments like routers. ALM is required to change topology structure dynamically because nodes often leave from and join multicast.

In addition, if NC applies to ALM, end hosts that take the place of routers are not limited. Thus, end hosts can do the encoding process, and the network does not require special equipment.

2.3 Liner Network Coding

Liner network coding (LNC) [3] is an encoding method that encodes packets with a liner vector. We use LNC as an encoding method. If LNC is used, let N packets be encoded and let intermediate nodes on path get encoding vector $\mathbf{c}=(c_1,c_2,\cdots,c_N)$ and do linear operation between this vector and arrived packets $p_i(i=1,2,\cdots,N)$. At this time, output packets p_{out} are then defined as follows:

$$p_{out}=c_1p_1+c_2p_2+\cdots+c_Np_N \quad (1)$$

where c_i is an element of a Galois field $GF(2^m)$ and packets are encoded as elements of a Galois field on a Galois field. In addition, these operations with multistage and p_{out} become input packets for the next intermediate node.

Let $\mathbf{x}=(x_1,x_2,\cdots,x_N)^T$ are N packets sent from a source node. The reception nodes need N different packets to extract original packets. Let N packets received by reception nodes $\mathbf{p}=(p_1,p_2,\cdots,p_N)^T$. p_i be expressed as formula (2) with coefficient

$$\tilde{\mathbf{c}}=(\tilde{c}_1p_1+\tilde{c}_2p_2+\cdots+\tilde{c}_Np_N) \quad (\tilde{c}_i,\ j\in GF(2^m),1\le i,\ j\le N)$$

$$p_i=\tilde{c}_{i,1}x_1+\tilde{c}_{i,2}x_2+\ldots+\tilde{c}_{i,N}x_N \quad (2)$$

Therefore, relationship between $\mathbf{p}$ and $\tilde{\mathbf{c}}_i$ is expressed as formulas (3), (4).

$$\mathbf{p}=\mathbf{C}\mathbf{x} \quad (3)$$

$$\mathbf{C}=\begin{pmatrix}\tilde{\mathbf{c}}_1\\ \tilde{\mathbf{c}}_2\\ \vdots\\ \tilde{\mathbf{c}}_N\end{pmatrix} \quad (4)$$

The original packet can be extracted with the presence of $\mathbf{C}^{-1}$, which is the inverse matrix of $\mathbf{C}$

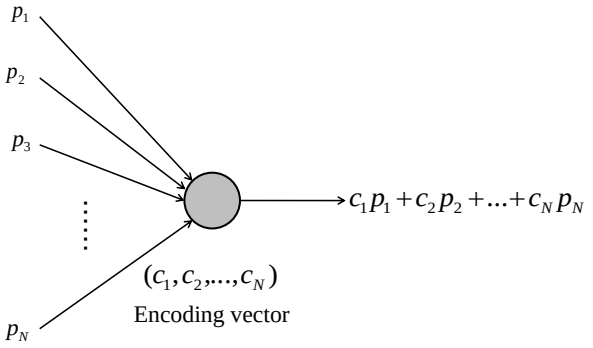


Figure 2: Liner network coding

3 PROPOSED METHOD

We developed a graph construction method for when NC is applied to ALM and a graph restoration method to resolve the frequently occurring problem of nodes leaving the ALM. We call these two proposed methods the join method and the leave method.

Let s be the source node, $t_i(i=1,2,\cdots,N_{rsv})$ be reception nodes, $u_j(j=1,2,\cdots,N_{\text{int}})$ be intermediate nodes, and class V be a multicast group composed of all nodes. Moreover, N_{rsv} is the number of reception nodes, N_{int} is number of intermediate nodes. Also, Max input link is $\Delta_{in}^{(x)}$ and Max output link is $\Delta_{out}^{(x)}$ with the presence of node x, Where $\Delta_{out}^{(x)}$ is expressed as $\Delta_{out}^{(x)}=B_x/B_{link}$. This B_x is the transmission band of the access link of x and B_{link} is the max transmission band of one logical link. Also, $\Delta_{in}^{(x)}=2$. Δ_x, which is the max connection of x, is given by the sum of max output link $\Delta_{out}^{(x)}$ and max input $\Delta_{in}^{(x)}$ link plus 1. Except Δ_x is limited by $\Delta_{\max}$ if it exceeds $\Delta_{\max}$ (formula (5)).

$$\Delta_x=\min\left\{\Delta_{out}^{(x)}+\Delta_{in}^{(x)}+1,\Delta_{\max}\right\} \quad (5)$$

In addition, it is assumed $\Delta_s=2$ on source node s.

3.1 Join Method

We use Nakai's method [4] and extend it by adding a mesh restoration method when nodes leave from multicast. This method has the three following steps when a node arrives in a multicast group.

a) First Step (forming mesh)

The node x that arrives in a multicast group quickly connects to a node belonging to the multicast group previously and is given multiple nodes addresses $V_x = \{v_{n1}, v_{n2}, \cdots, v_{n|vx|} \mid v_{nk} \in V\}$ in multicast. Then x connects to a high priority node in the following step and connects to nodes up to Δ_x. We define the lower delay as higher priority. If more than one node has the same delay, we define the wider bandwidth and higher priority. And $|V_x| > \Delta_x$.

(1) $P = V_x$.
(2) Pick the highest priority node y from x in P and $P = P - \{y\}$.
(3) Node x sends a connection request message to y. Node y sends a connection acceptance message if the number of neighboring nodes of y is less than Δ_y and sends a connection rejection message if the number of neighboring nodes of y is more than Δ_y. Node x adds node y to its own neighboring node list when it receives a connection acceptance message.
(4) Turn back to (2) if $P \neq \phi$ and the number of neighboring nodes of x is less than Δ_x.

Each node in the multicast group regularly exchanges information it know and keeps all node addresses in the multicast group. The node to which a new attending node connects is not optimal for delay. Thus topology is optimized regularly. This optimization method is the extended Nakai method with the mesh restoration method. Specifically, each node n selects node m from non-neighboring nodes in random order regularly.

(1) Add m to n's neighboring node list if the number of n's neighboring nodes is less than Δ_n and the number of m's neighboring nodes is less than Δ_m.
(2) Other than (1), if the lower priority neighbor node than m is exists, exchange it for m.

b) Second Step (making first path)

A construct tree including all reception nodes by using a distance vector (DV) algorithm [5], the metric of which is delay on mesh, is formed in the first step. If some paths have the same delay, select a wider bandwidth path.

c) Third Step (making second path)

Construct a second path that is node-disjoint from the first path to each reception node with the following steps.

(1) $i = 1$
(2) Consider the link from each node to its neighboring nodes as directed in graphs and do the following operation for all nodes. Node x 's output link is deleted except when used in the first and second paths if x 's output link outdegree is $(x) \geq \Delta_{out}^{(x)}$. In addition, node x 's input link is deleted except when used in the first and second paths if x 's input link indegree is $(x) \geq \Delta_{in}^{(x)}$.
(3) The delay of links used in the first path from source node s to reception node t_i is set to infinity, and that used in first path and second path except path from source node s to reception node t_i is multiplied α $(0 \leq \alpha < 1)$.
(4) The second path from source node s to reception node s is set by DV algorithm the metric of which is delayed.
(5) The delay which be set in (3) is undone. Then, turn back to step (2) on $i = i + 1$ if $i \leq N_{rcv}$.

Figure 3 shows an example of the join method. Figure 3(a) is mesh topology formed in the first step. The thick lines in figure 3(b) are the first path. Figure 3(c) shows the second path to t_1, and x-marks on links express that the delays of these links are set to infinite. The extra-wide lines express that the delay of these links are multiplied α. Figure 3(d) show second path to t_2 made after making second path to t_1.

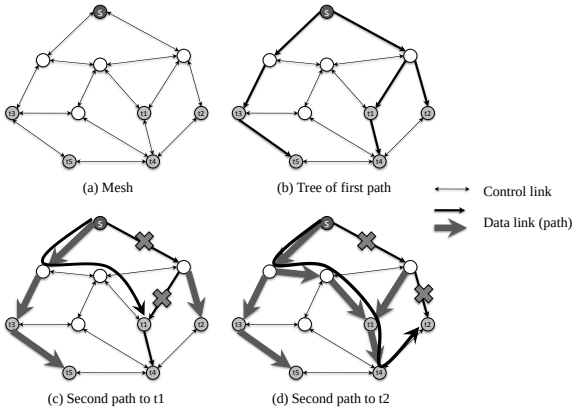


Figure 3: Join method

3.2 Leave Method

When a node leaves the multicast group normally, the node can notify others of its will to leave before it leaves. As mentioned in 2.3, we use LMC as an encoding method. Thus, if the encoding vector in

intermediate nodes often changes when nodes leave, reception nodes may not extract the coding packet. Therefore, the encoding vector in intermediate nodes must not change when paths are restored.

We normally use the restoration method when a node leaves from multicast group as follows. We let upstream node be a parent node and downstream node be a child node focusing on the data stream. In addition, all nodes have alternate nodes that can substitute for them when they leave, except when source node and end nodes are in a tree. The nodes chose an alternate node with the lowest delay from its child nodes (Figure 4(a)).

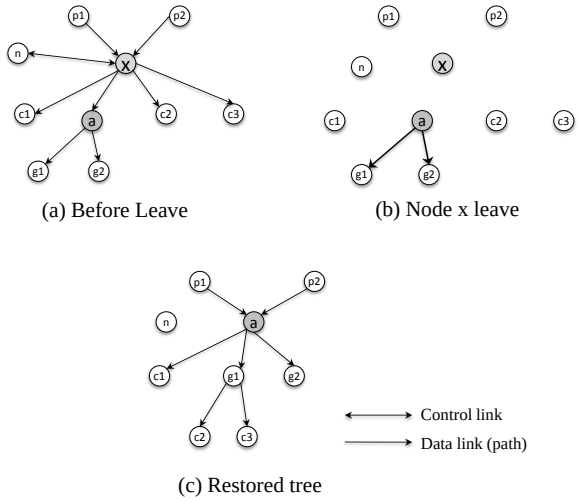
(1) Leave node x sends leave message to its neighboring nodes. The leave message contains its alternate node's information, and information of parent nodes p1 and p2 and its encoding vector.

(2) The nodes that receive the leave message cut off the link to node x (Figure 4(b)).

(3) Alternate node a connects to x 's parent nodes p1 and p2. x 's parent nodes reject connection requests from the other nodes until they receive a connection request from node a .

(4) Child nodes $c_i (i = 1,2,\cdots)$ except node a send connection requests to node a by using leave message information. Node a sends a connection acceptance message if its output link outdegree (a) is less than max output link $\Delta_{out}^{(a)}$. Otherwise it returns the connection rejection message and information of its child node $g_i (i = 1,2,\cdots)$. If c_i gets a connection rejection message, it sends a connection request to the node that has the lowest delay in g_i continuously.

Therefore, c_i reconnects node a or a 's children nodes seriatim (Figure 4(c)). The paths are restored like this. After path restoration, a encodes packets received by encoding vector of leave node contained in the leave message. In this way, nodes under the a node receive the same packets as before after node x leaves. Furthermore, neighboring nodes n of leave node x are restored and connected to optimum nodes by the mesh optimization method mentioned in 3.1 a).



(a) Before Leave (b) Node x leave

(c) Restored tree

Figure 4: Leave method

4 EXPERIMENT TO EVALUATE

We conducted a simulation to evaluate the proposed methods' performance and verify the leave method. We use NS-2 [6] as simulator and simulate on simple physical topology.

Physical topology is designed as in Figure 5 with 4 routers and 12 end nodes.

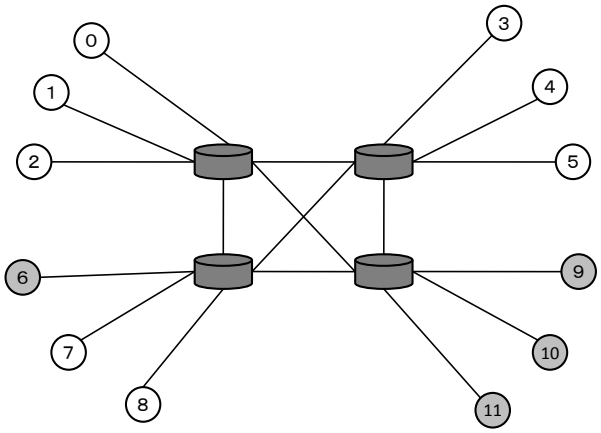


Figure 5: Simulation model

Table 1 shows the parameters used in the simulation.

Table 1: Simulation parameters

Band between routers	2Mbps
Delay between routers	10msec
Band between routers and node	Uniform distribution in [2,10] Mbps
Delay between routers and nodes	Uniform distribution in [0,1] msec
a	0.5
Optimize Interval	1000msec

We use $a = 0.5$, which Nakai's method [3] says is the optimum a used in path construction. Gray end nodes in Figure 5 are reception nodes in ALM. Therefore, of the total, 1 is a source node, 4 are reception nodes, and 7 are intermediate nodes.

We simulate making the first path, second path, data packet transfer, and node leaving. The data packet transfer rate is 32kbps.

In this case, we weigh up the proposed and non-coded ALM with only first path.

When a node leaves, we measure data packet that the leaving node could not get to get data packet again by tree restoration. In addition, we weigh up the number of control message non-leaving cases and leaving cases.

5 RESULTS AND CONSIDERATIONS

Figure 6 shows an example overlay network shown constructed by mesh optimizing. First and second paths from source node to each reception node (6, 9, 10, and 11) are also constructed.

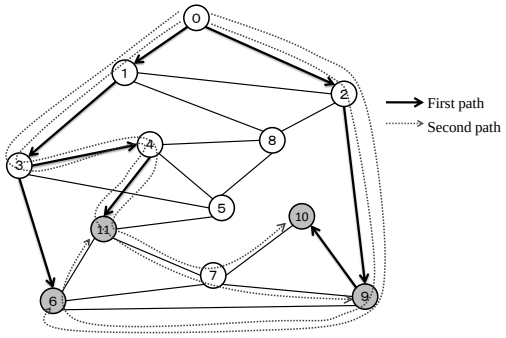


Figure 6: Paths constructed by simulation

Figure 7 evaluates performance, throughput, and delay of data packets. In this graph, normalized link stress means the average traffic on all link normalized by its throughput. Normalized link stress decreases 13% from non-coded case. However, delay increases 48% from non-coded ALM.

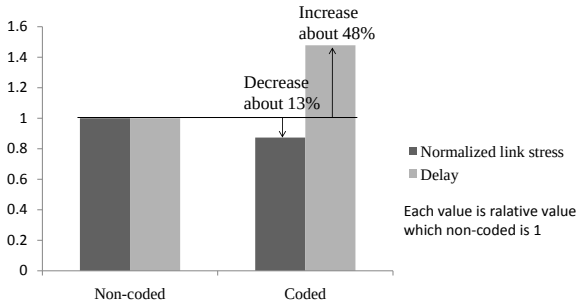


Figure7: Normalized link stress and delay

Normalized link stress is attributed to the effect of decreasing network load. However, this is relatively ineffective because topology is small and few middle links overlap. The most suitable path for delay is the first path, which is superior to the second. Despite this, high delay path is given most priority in the nature of NC.

Node 3 leaves in Figure 6. The tree is restored in Figure 8 by the Leave method. In this case, the alternate node of node 3 is node 4, so node 4 connects to node 1, which is the parent node of the leave node. Also, node 6 connects to node 4, which is the alternate node of the leave node.

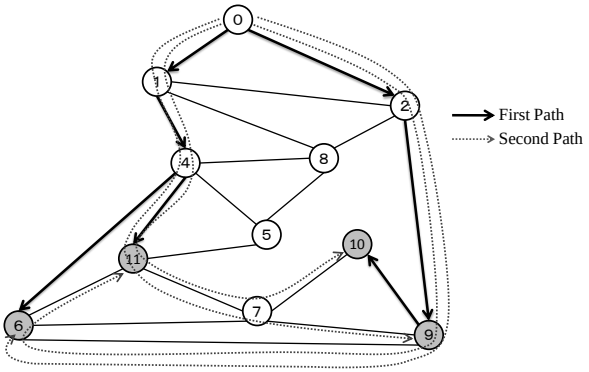


Figure 8: Restored path after node 3 leaves

At this time, child nodes of the leaving node are affected by it leaving. Table 2 shows the times that node 4 and node 6, which are child nodes of node 3, take and packets that these nodes cannot receive. Times to restore are about 20msec for each node. These are very short compared with the transfer rate. In fact, the number of packets that these nodes cannot receive during tree divide is only one. This effect can be tackled by data buffering.

Table 2: Effect for child nodes of leaving node

	Time to restore	Number of unreceivable packets
Node 4	18msec	1
Node 6	20msec	1

The number of control packet in all topology is shown in Figure 9, which compares the leaving case with the non-leaving case. Increases in control packets are very few, and there is no affect when one node leaves from the tree.

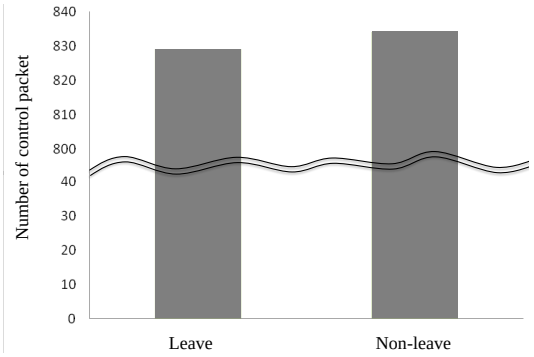


Figure 9: Comparison of number of control packets

6 CONCLUSION

In this research, we focused on solving the network load increase problem in ALM. We extended an existing method to solve the path construction problem and proposed applicable method that can restore path normally when a node leaves. Then we verified practically by simulation.

Next, we will conduct an experiment on the basis of requirements and consider the applicable cases. In addition, we will conduct the experiment in a large-scale environment and elucidate performance of the proposed method in this environment.

Additionally, we will propose a path restoration method for when a node leaving on its own cannot leave normally and send messages to neighbors. Then, we propose to switch coded and non-coded ALM dynamically so that solve problem of delay increasing. Therefore we will increase the feasibility of ALM using NC.

REFERENCES

[1] Y. Chu, S. G. Rao, S. Seshan, and H. Zhang, A Case for End System Multicast, IEEE Journal on Selected Areas in Communications, Vol. 20, No. 8, pp. 1456-1471, (2002).

[2] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, Network Information Flow, IEEE Transactions on Information Theory, Vol. 46, No. 4, pp. 1204-1216, (2000).

[3] S.-Y. R. Li, R. W. Yeung, and N. Cai, Linear Network Coding, IEEE Transactions on Information Theory, Vol. 49, No. 2, pp.371-381, (2003).

[4] Takayuki NAKAI, Taku NOGUCHI, Takahiro MATSUDA and Tetuya TAKINE, Graph Construction for Application Level Multicast Using Network Coding, IEICE Technical Report, Vol. 107, No. 525, IN2007-160, pp. 7-12, (2008).

[5] Z. Wang and J. Crowcroft, Quality of Service Routing for Supporting multimedia Application, IEEE Journal on Selected Areas in Communications, Vol. 14, No. 7, pp.1228-1234, (1996).

[6] The Network Simulator 2 (ns-2), http://www.isi.edu/nsnam/ns/.

Evaluation of Mitigation to Bursty Packets by a TCP proxy in a Wired and Wireless Network

Toshihiro Shikama†

†Fukui University of Technology, Japan
shikama@fukui-ut.ac.jp

Abstract - This paper investigates a TCP proxy that splits a TCP connection into two parts consisting of a wireless link and a wired network. The TCP proxy is effective for improving TCP performance in such a heterogeneous network including a wireless link, on which transmission errors occur. This paper describes how the TCP proxy produces large size forward data due to a packet loss on the wireless link. It also identifies a new problem that the output traffic from the TCP proxy becomes bursty due to the large size forward data. To mitigate bursts of packets by the TCP proxy this paper proposes a scheme that the proxy performs a pacing function, which places a gap between two consecutive packets. Since this function is performed in cooperation with the flow control between two TCP connections, the scheme has an advantage that it requires a small amount of forwarding buffers for the pacing. Simulation results using ns-2 show that bursty packets are produced by the conventional TCP proxy and the pacing function suppresses them. The results also show that throughput is improved by the proposed scheme, while the throughput of the conventional TCP proxy suffers from packet losses in the wired network due to the bursts of packets.

Keywords: TCP proxy, PEP, burst, pacing, reassembling

1 Introduction

Transmission errors have to be considered in IP networks that employ wireless links. Although TCP is mainly employed end-to-end in the Internet, it is well understood that TCP cannot achieve sufficient throughput in the environment where packets are lost due to transmission errors. TCP assumes that losses of packets are derived from a buffer overflow at a forwarding node; it invokes the congestion control to reduce traffic. In the case where losses of packets are caused by transmission errors on a wireless link, the congestion control is performed unnecessarily; a degradation of TCP throughput occurs.

One approach to mitigate this problem is to employ a TCP proxy called PEP (Performance Enhancing Proxy) that terminates a TCP connection from a source terminal and establishes another TCP connection to the destination terminal [1]. It forwards receive data from one TCP connection to another. As the large round-trip delay causes low throughput and the PEP makes the delay short by splitting a TCP connection, the throughput is improved. Although termination of a TCP connection by the PEP has a significant impact on the characteristics of an IP flow, a study on this aspect has not been done so far. This paper focuses on the property of output traffic from the PEP and identifies a problem concerning bursts of packets due to the termination of a TCP connection. It also proposes a mitigation scheme for the problem and evaluates its effectiveness by simulations using ns-2.

This paper assumes that a PEP is placed in a node which connects a wireless link to a wired network. In the following description, we introduce the term “a wireless section” to refer to the wireless link. We also call the wired network “a wired section”. Figure 1 illustrates an example of an error recovery sequence by the PEP placed between the wireless section and the wired section. When a loss of a packet occurs on the wireless section, an error recovery by a retransmission of associated segment is performed by TCP over the wireless section. The figure assumes that the fast retransmit is invoked after three duplicate ACKs. If a receiving side of TCP in the PEP accepts out-of-order segments after the lost one, it retains them in its receive buffer to reassemble them. When the lost segment is retransmitted and received correctly, the reassembling of data is completed; the whole data consisting of the lost segment as well as the buffered segments is forwarded to the next TCP connection over the wired section. In this paper, we call this data “forward data”. Since the size of the forward data might be up to the window size of TCP over the wireless section, a large number of segments might be generated at the same time. In Figure 1, a burst of packets is immediately sent by the PEP after the arrival of the retransmitted segment. The burst of packets should be avoided, since it may cause a buffer overflow at a node forwarding these packets on the wired section.

With regard to this busty packets from TCP, RFC 2581 specifies that TCP should employ a slow-start after a silent period of more than one RTO (Retransmission Time Out) [2]. In the normal case, when an application program issues a send request for large size data, the transmission of segments begins with the slow-start and enters the ACK clocking state after some transient period [3]. The burst of packets occurs during the period of the slow-start only. The maximum size of a burst by the slow-start is about a half of the congestion window, while the size of a burst by the PEP is up to the congestion window. This means that the maximum size of a burst by the PEP is about twice as large as that by the slow-start.

In Figure 1, there is an idle period over the wired section from the loss of the segment to the reception of the retransmitted segment. However, the duration of this period is comparable with one round-trip time of the wireless section, which is much less than the RTO period. Therefore the slow-start is not invoked in this case; a burst of packets derived from the large size forward data is likely fed into the wired section.

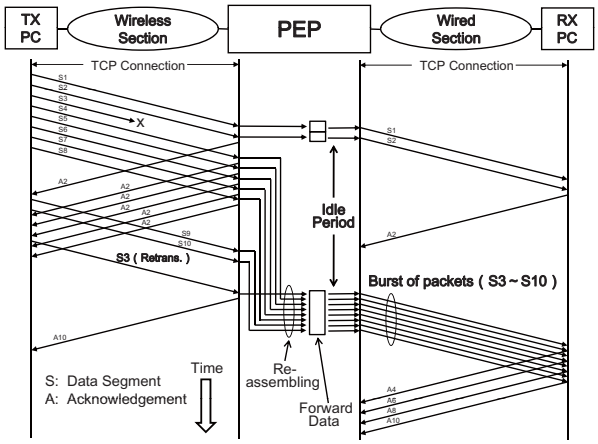


Figure 1: Data forwarding by a PEP and its output traffic.

In the case of the conventional end-to-end TCP, a burst of packets of which size is equal to the TCP window may occur, if two send requests of large size data are issued within the interval of less than one RTO. The second send request may cause such large burst, since the slow start is not invoked on this condition. However, in the case of conventional applications based on the client-server model, a send request of large size data occurs after the previous transfer of large size data to a destination client has been completed and the client issues the next data request. As the interval between two send requests is usually larger than the RTO period, the large size burst issued by TCP is not a significant problem for the conventional end-to-end TCP.

The rest of this paper is organized as follows: Section 2 discusses related work concerning PEP. Section 3 discusses the condition of burst generation and proposes a scheme to mitigate bursts of packets. Section 4 describes the simulation models and associated assumptions. Section 5 describes simulation results and discussion. Finally our conclusion is presented in Section 6.

2 Related work

Concerning the problem of the poor performance by TCP for a network including a wireless link, improvements of the congestion control done by TCP are studied. A typical version of TCP is Westwood [4], which estimates an available bandwidth. When it detects a packet loss, it identities the cause of the loss by checking whether the available bandwidth is larger than the actual transmission rate. If this is the case, the congestion control is not invoked. This approach of improving TCP has a problem that TCP in an existing PC has to be changed.

Another approach to improve the TCP performance in a wireless network is to employ local retransmissions by Link Layer protocol over the wireless section [5]. Since the Link Layer protocol hides from TCP all packet losses over the wireless section, the performance of TCP is improved. However, the modification of the protocol stack in an existing PC is needed. This approach also has the problem of spurious timeouts in TCP. Snoop [6] is a kind of local retransmission scheme, which uses duplicate TCP ACKs to invoke a retransmission of a lost packet. This scheme has an advantage of no modification of an existing PC. However, since it merely performs retransmissions of lost packets, packets arrive at the destination out-of-order. It requires a mechanism to avoid unnecessary duplicate ACKs, which is hard to implement if TCP employs SACKs.

PEPs are traditionally employed in satellite networks where a propagation delay is large. However, it is also effective to improve throughput of terrestrial radio networks of which transmission rate is increasing rapidly [8]. RFC3135 surveys various PEP architectures and their effects on the performance as well as reliability of a system. It also describes details of the controversial problem that PEP cannot keep end-to-end semantics of TCP acknowledgement. It refers to the possibility that PEP may send a burst of data segments due to ACK handling done by PEP. It also describes the scheme that places a gap between ACKs to suppress the bursts. However, it never refers to the bursts of segments by the reassembling function done by PEP.

The PEP architectures include a scheme, in which PEP does not terminate a TCP connection, but it returns an early ACK to the sender on behalf of the destination. As PEP does not terminate a TCP connection, bursts of segments due to the reassembling never occur in this scheme [7]. Since this scheme has to keep copies of forwarded segments in provision for segment losses for which PEP has sent ACKs, it has problems that complex retransmission procedures as well as the buffer management are required.

Concerning bursts of packets sent by TCP for the case where TCP is employed end-to-end, bursts of packets issued by TCP are studied and classified into micro-bursts and macro-bursts [9]. However, the cause of bursts by the PEP described in the previous section is different from mechanisms described in the literature. A natural way to mitigate the bursts of packets is to suppress the peak rate of packets at the output of TCP. In a Linux environment, a simple way is to employ TBF (Token Bucket Filter) to shape the output traffic from TCP. However, if there are multiple TCP connections, it is hard for this scheme to suppress the peak rate of each TCP connection.

A scheme that suppresses the peak rate of packets precisely at the output of TCP has been proposed in Linux [10]. This scheme employs a special PAUSE frame to place a gap between two consecutive packets. However, if there are multiple TCP connections, it is hard for this scheme to suppress the peak rate of each TCP connection. In addition to this problem, although this scheme needs a large number of buffers to perform the packet pacing, it is hard to predict the required number of buffers. Packet losses due to internal buffer overflows may occur at the pacing, if there are a large number of TCP connections.

3 The condition of burst generation and a mitigation scheme for bursts

3.1 The condition on bursts of packets by TCP

Since a sending side of TCP stores send data in its buffer and it performs the window flow control, a send request for large data may not directly cause a burst of packets by TCP.

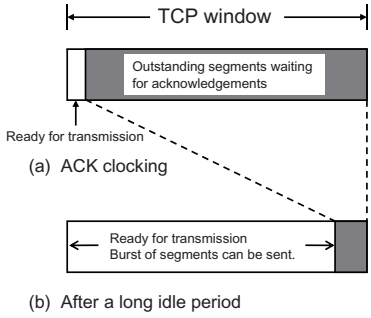


Figure 2: The condition on the occurrence of packet bursts.

However, the pacing at the input of TCP can be effective for the following reasons. When TCP continues the transfer of data in a steady state, ACK clocking is effective. In this case bursts of packets are not generated, even though a send request for large data is issued. Figure 2 (a) shows the window of this condition. After TCP sends data segments of TCP window size, it waits for an ACK from the receiving side of TCP; it cannot send further segments even though a new request for send is issued. Transmission of segments is performed, only when a new ACK arrives; TCP cannot be disturbed by the arrival of a new send request. In this condition, bursts of packets never occur.

Bursts of segments can be sent by TCP, when the window is open largely as shown in Figure 2 (b). Segments up to the open window size can be immediately send by TCP, when a send request for large data is invoked. From Figure 1, as mentioned before, there is an idle period over the wired section from the loss of a packet to the reception of the retransmitted packet. During this period, since a number of ACKs are received by the sending side of TCP of the wired section, a substantial number of outstanding segments are acknowledged; the TCP window opens largely, before a send request for the large size forward data is issued. Therefore, a burst of segments is sent by TCP.

3.2 A proposed scheme

This paper proposes a scheme that performs the pacing at the input of TCP as shown in Figure 3. The pacing function is performed by the application program that forwards data from TCP. This function segments the large size forward data into multiple data chunks of which size is MSS over the wired section. It issues a send request for each data chunk to TCP at a specific rate. This scheme has the following advantages.

- Pacing on each TCP connection can be performed independently.
- Pacing can be performed in cooperation with the flow control between two TCP connections. If large size data is forwarded from the wireless section to the wired section, the pacing function is able to stop receiving further data from the wireless section until the send requests for all chunks of the large size forward data have been completed. Because of this flow control, a packet loss due to a buffer limit never occurs at the PEP.

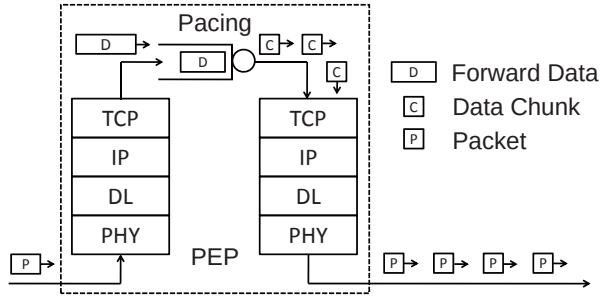


Figure 3: Relation between forward data and bursts of packets.

3.3 A pacing rate

When the pacing is employed, a problem is how to determine the suitable packet rate after the pacing. In this paper, we simply apply the packet rate derived from the transmission rate of the wireless section, assuming that the rate of the wireless section is smaller than the bottleneck rate of the wired section. There are possibilities of sophisticated ways to determine the pacing rate, for example, estimation of end-to-end throughput over the wired section. These are left for further study.

4 Simulation Models

This paper evaluates output traffic from the PEP and effectiveness of the proposed scheme by simulations using ns-2 [11]. We compare the following three schemes.

- Scheme A: PEP with pacing (proposed scheme)
- Scheme B: PEP without pacing (conventional PEP)
- Scheme C: No PEP (end-to-end TCP)

Figure 4 shows the simulation model for the proposed scheme where the PEP with pacing is employed, while Figure 5 shows the simulation model for the conventional PEP where the pacing is not employed. NewReno TCP is employed for both the wireless and wired sections. Node n1 represents the PEP that terminates a TCP connection over the wireless section (from node n0 to node n1) and forwards received data to the next TCP connection over the wired section (from node n1 to node n3). Since original ns-2 does not have the function of delivering received data to the upper application layer, we added this function to a TCP sink in ns-2. This is indicated as "TCP Sink+" in Figures 4 and 5. We measured the throughput of these schemes as well as the distributions of the number of packets included in a burst sent by the PEP.

The bandwidth of the wireless section is assumed to be 5 Mbit/s, while the PEP is connected to node n2 with the rate of 100 Mbit/s. Node n2 is connected to destination node n3 with the rate of 10 Mbit/s, which is assumed to be the bottleneck bandwidth B_W over the wired section. Since there is a difference of packet transmission rate between input and output in node n2, bursts of packets may be queued at this node. Moreover, we assume insufficient buffering at a node connected to the bottleneck link. The number of buffers in node n2 is

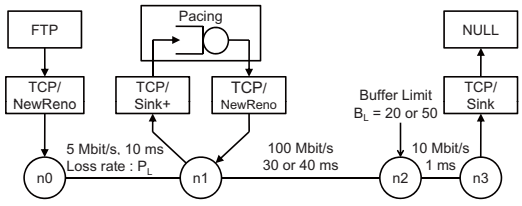


Figure 4: Simulation model for Scheme A (PEP with pacing).

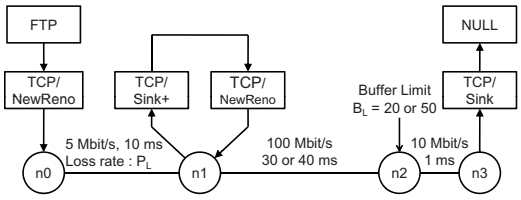


Figure 5: Simulation model for Scheme B (PEP without pacing).

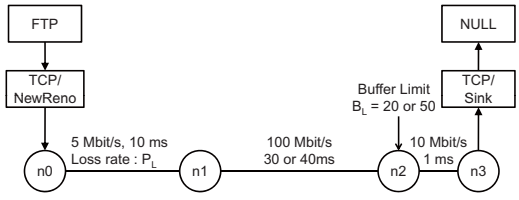


Figure 6: Simulation model for Scheme C (No PEP).

limited; packets may be lost when a burst of packets is fed into this node. In order to confirm the improvement of the total throughput by the PEP we also simulate the case where a PEP is not employed. Figure 6 shows the model of scheme C (No PEP), where an error recovery is performed by TCP end-to-end. Table 1 summarizes the simulation parameters.

5 Simulation results and discussion

5.1 The distributions of the forward data sizes and the number of packets in a burst, where node n2 has sufficient buffering

Figure 7 shows the distributions of both the forward data sizes and the number of packets in a burst, where node n2 has enough number of buffers and the pacing is not performed. The packet loss rate P_L is selected to 0.001 and 0.01. As the original size of forward data is measured in bytes, the values of data size indicated in the figure are divided by MSS (Maximum Segment Size: 1460 B) to compare with the number of packets in a burst. In this figure rectangle boxes (Data size) represent the distribution of the forward data sizes, while black dots (Burst size) show the distribution of the number of packets in a burst, which are sent by TCP.

The maximum size of forward data is about 43, which is consistent with MSS and the size of TCP window (64 KB) over the wireless section. We can observe that the distributions of the forward data size and the number of packets in a burst are well consistent. This means that the PEP sends bursts of packets corresponding to the forward data sizes. In the case of the small packet loss rate (P_L = 0.001), although

Table 1: Simulation parameters

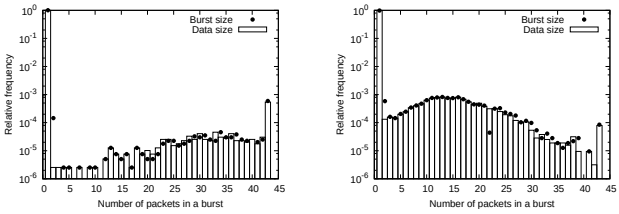
Rate of the wireless section	5 Mbit/s
Delay of the wireless section	10 ms
TCP version for the wireless section	NewReno
TCP window for the wireless section	64 KB
Generation of packet losses	random
Packet loss rate: P_L	0.0001, 0.0002, ... , 0.05, 0.1
TCP version for the wired section	NewReno
TCP window size for the wired section	128 KB
Delay of the wired section: D_W	30 ms, 40 ms
Window size of end-to-end TCP	128 KB
The number of TCP connections	1
Rate of the wired section (PEP to n2)	100 Mbit/s
Interface rate from node n2 to n3: B_W	10 Mbit/s
The number of buffers at node n2: B_L	20, 50 packets
Simulation time	1000 sec
The number of simulation runs	10

the frequency of burst occurrence is generally small, there is a trend that large size bursts occur compared with the case of the large packet loss rate (P_L = 0.01). The reason for this trend is as follows. When TCP detects a loss of a segment by duplicate ACKs, it retransmits the lost segment and decreases its congestion window (cwnd) in half. In the case of the large packet loss rate (P_L = 0.01), the decrease in cwnd occurs more frequently compared with the case of the small packet loss rate (P_L = 0.001). As the size of cwnd becomes small on the average and TCP window is limited by cwnd, the size of forward data in the case of large packet loss rate (P_L = 0.01) tends to be small.

Figure 8 shows the same distributions as Figure 7, except that the pacing is performed by the PEP. Although the distribution of the forward data sizes represented by rectangle boxes (Data size) is the same as Figure 7, the distribution of the number of packets in a burst represented by black dots (Burst size) is completely different. The number of packets in a burst is suppressed to up to 2. This is due to the effect of the pacing. The reason of two packets in a burst after the pacing is that TCP employs the delayed ACK.

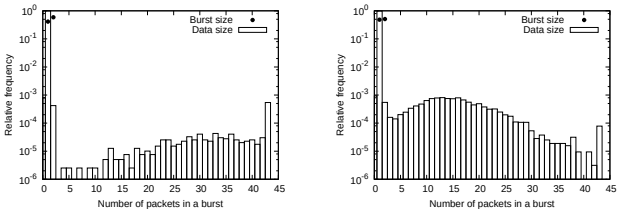
5.2 Throughput for the case where the number of buffers at node n2 is limited

Figure 9 shows the relation between the packet loss rate over the wireless section and the total throughput, where the number of buffers at node n2 is limited to 20. Throughput generally decreases as the packet loss rate becomes large. However, the decrease in throughput is suppressed significantly by the PEP with the pacing; scheme A achieves the best performance among three schemes. When the packet loss rate is small, the throughput of the conventional PEP (scheme B) is smaller than scheme C, in which an error recovery is performed end-to-end. The reason why the throughput of scheme B is inferior to scheme C is that packet losses occur at node n2 due to bursts of packets by the conventional PEP.



$P_L = 0.001$ $P_L = 0.01$

Figure 7: The distribution of the forward data sizes and the number of packets in a burst, where pacing is not performed.



$P_L = 0.001$ $P_L = 0.01$

Figure 8: The distribution of the forward data sizes and the number of packets in a burst, where pacing is performed.

Figure 10 shows the relation between the packet loss rate over the wireless section and that over the wired section. Large packet loss rate is observed in scheme B, since packet losses occur due to buffer overflows at node n2.The packet loss rate over the wired section becomes larger than or comparable with that of the wireless section, when the packet loss rate of the wireless section is less than 0.005. We cannot observe any packet loss in schemes A and C. Since there is no packet loss and RTT of the wireless section becomes small due to the PEP architecture, the performance of scheme A is the best. It is clear that the effect of the pacing is outstanding.

5.3 Throughput for the case where the number of buffers at node n2 becomes large

Figure 11 represents the total throughput for the case where the number of buffers at node n2 is increased from 20 to 50. Although the throughput of scheme B is much improved, its value is still lower than that of scheme A for cases of small packet loss rates (less than 0.005). When the packet loss rate is 0.005, the throughput of scheme B takes larger value than the throughput values of lower packet loss rates.

Figure 12 shows the relation between packet loss rate of the wireless section and that of the wired section. Due to the increase in the number of buffers, the packet loss rates of scheme B become smaller than those of Figure 10. Although the packet loss rate over the wired section becomes small in comparison with the case of Figure 10, its value is still larger than that of the wireless section for the cases of the packet loss rates less than or equal to 0.001. Similar to the case of Figure 9, this explains the degradation of the throughput when the packet loss rate of the wireless section is small. As there is no packet loss in schemes A and C, the performance of these schemes is the same as the case of Figure 9.

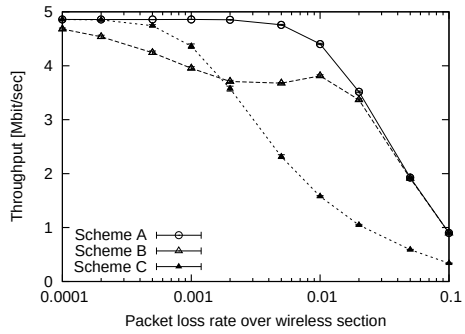


Figure 9: Packet loss rate vs. throughput, D_W = 30ms, B_L = 20, B_W = 10 M bit/s.

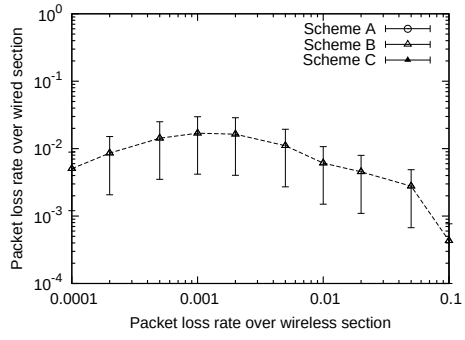


Figure 10: Packet loss rate over wireless section vs. that over wired section, D_W = 30ms, B_L = 20, B_W = 10 M bit/s.

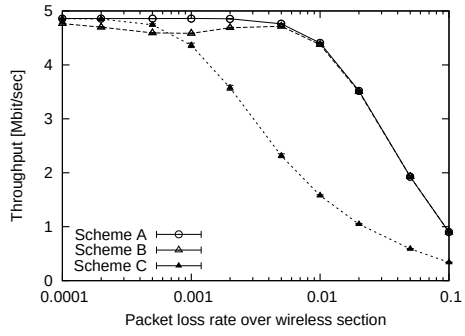


Figure 11: Packet loss rate vs. throughput, D_W = 30ms, B_L = 50, B_W = 10 M bit/s.

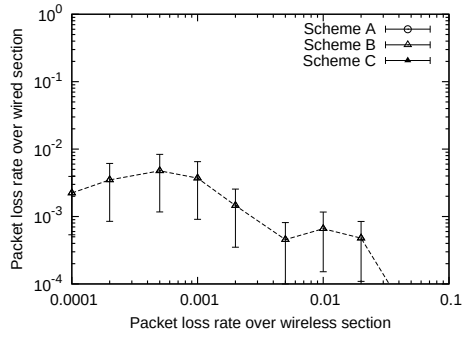


Figure 12: Packet loss rate over wireless section vs. that over wired section, D_W = 30ms, B_L = 50, B_W = 10 M bit/s.

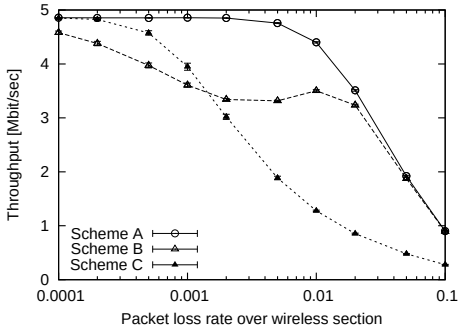


Figure 13: Packet loss rate vs. throughput, D_W = 40ms, B_L = 20, B_W = 10 M bit/s.

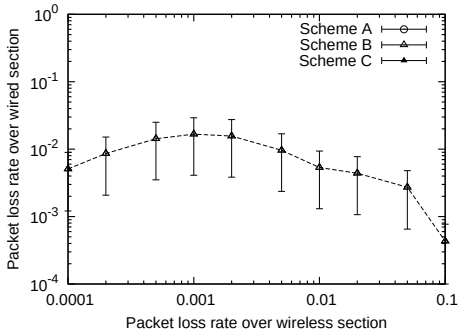


Figure 14: Packet loss rate over wireless section vs. that over wired section, D_W = 40ms, B_L = 20, B_W = 10 M bit/s.

5.4 Throughput for the case where the delay of wired section becomes large

Figure 13 shows the relation between the packet loss rate over the wireless section and the total throughput, where the delay of the wired section is 40ms, which is 10ms larger than the case of Figure 9. The throughput of scheme B becomes worse while the throughput of scheme A is the same as the case of Figure 9. In the case of scheme A, although the maximum throughput of the wired section becomes small due to the increase in the delay of the wired section, its value is still larger than the throughput of the wireless section. Then, the total throughput scheme A is the same as the case of Figure 9. Since the increase in the delay of the wired section is relatively small compared with with the total end-to-end delay, the throughput of scheme C is also almost the same as Figure 9.

Figure 14 shows the relation between the packet loss rate over the wireless section and that over the wired section. This figure is almost the same as Figure 10. As the RTT of the wired section increases, the effect of packet losses to the throughput becomes large. The reason is that losses of packets cause decrease in the congestion window (cwnd), which leads to the decrease in throughput. This indicates that the effect of bursts of packets by the conventional PEP becomes serious when the delay of the wired section becomes large.

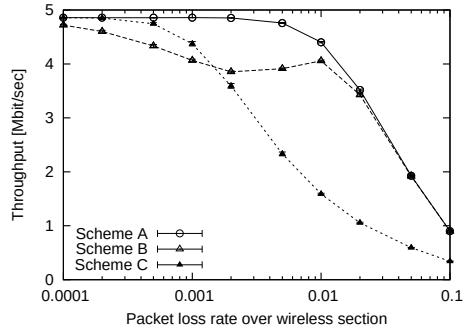


Figure 15: Packet loss rate vs. throughput, D_W = 30ms, B_L = 20, B_W = 20 M bit/s.

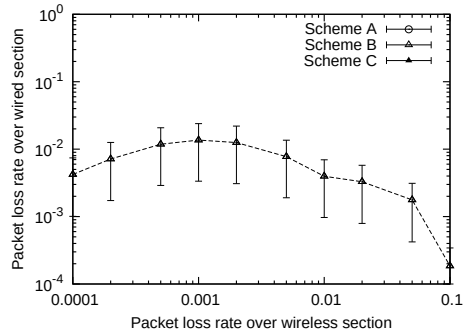


Figure 16: Packet loss rate over wireless section vs. that over wired section, D_W = 30ms, B_L = 20, B_W = 20 M bit/s.

5.5 Throughput for the case where the bandwidth of the bottleneck link becomes large

All the results presented above are taken on the condition that the bandwidth of the bottleneck link B_W is 10 Mbit/s. Figure 15 shows the relation between the total throughput and the packet loss rate over the wireless section, when the bandwidth B_W is doubled (20 Mbit/s) and the number of buffers at node n2 is 20. The throughput of scheme A and C is the same as the case of Figure 9, since there is no packet loss in Figure 15 and Figure 9. We can observe improvement of throughput of scheme B compared with Figure 9. However, the throughput of scheme B is still the smallest among the three schemes for the cases of small packet loss rates.

Figure 16 shows the relation between the packet loss rate of the wireless section and that of the wired section. Although the packet loss rate of the wired section is improved, its value is still large. We cannot observe significant effectiveness of the increase in the bottleneck bandwidth. This shows that the increase in the bandwidth of the bottleneck link has small impact on suppressing packet losses due to a large burst of packets.

6 Conclusion

This paper has investigated the architecture of a TCP proxy (PEP), where the PEP forwards receive data from one TCP connection on the wireless section to another TCP connection on the wired section. We have focused on the output traffic from the PEP. When losses of packets occur due to trans-

mission errors, we have identified a problem that the output traffic from the PEP becomes bursty due to the reassembling function done by the receiving side of TCP over the wireless section. To mitigate the bursts of packets we proposed a pacing function at the PEP. This scheme has advantages of no packet loss with a limited amount of buffers and independent pacing for each TCP connection. We simulated the case where the output traffic from the PEP is forwarded by a node that is connected to a bottleneck link in the wired section and the node has insufficient number of buffers compared with the window size of TCP. The simulation results have shown that the performance of the PEP with pacing is improved significantly. We have also observed that the total throughput of the conventional PEP becomes worse than that of end-to-end TCP for the cases of small packet loss rates. As the rate of wireless links is increasing, a TCP proxy will be important to attain high throughput. Accordingly, a burst of packets by the TCP proxy will be serious and its mitigation will be significant.

Acknowledgment

This work has been supported by KAKENHI (20500079) and Special Research Grant in Aids of Fukui University of Technology.

REFERENCES

[1] J. Border, M. Kojo, J. Griner, G. Montenegro, and Z. Shelby, “Performance Enhancing Proxies Intended to Mitigate Link-Related Degradations,” IETF, RFC3135, June 2001.

[2] M. Allman, V. Paxson, and W. Tevens, “TCP Congestion Control,” RFC 2581, April 1999.

[3] W. R. Stevens, “TCP/IP Illustrated, Volume 1: The Protocols,” Addison-Wesley, 1994.

[4] Casetti, M. Gerla, S. Mascolo, M. Y. Sanadidi, and R. Wang, “TCP westwood: Bandwidth estimation for enhanced transport over wireless links,” ACM MOBICOM 2001, pp.287-297, July 2001.

[5] G. Fairhurst and L. Wood, “Advice to Link Designers on Link Automatic Repeat reQuest (ARQ),” IETF, RFC3366, August 2002.

[6] H. Balakrishnan, S. Seshan, E. Amir, and R. Katz, “Improving TCP Performance over Wireless Networks,” 1st ACM International Conference on Mobile Computing and Networking (Mobicom), Nov. 1995.

[7] D. Dutta and Y. Zhang, “An Active Proxy Based Architecture for TCP in Heterogeneous Variable Bandwidth Network,” IEEE GLOBECOM 2001, pp.2316-2320, Nov. 2001.

[8] M. Meyer, J. Sachs, and M. Holzke, “Performance evaluation of a TCP proxy in WCDMA networks,” IEEE Wireless Commun., vol.10, no.5, pp.70-79, Oct. 2003.

[9] M. Allman and E. Blanton, “Notes on Burst Mitigation for Transport Protocols,” ACM Computer Communications Review, Vol, 35(2), pp.53-60, April 2005.

[10] R. Takano, T. Kudo, Y. Kodama, M. Matsuda, F. Okazaki, and Y Ishikawa, “Improving TCP Performance by Using Precise Software Pacing Method,” IEICE Technical Report, NS2005-157(2006-1).

[11] The Network Simulator - ns-2, http://www.isi.edu/nsnam/ns/, 2010.

A Survey of Reprogramming Systems in Wireless Sensor Networks

Aoi Hashizume†, Hiroshi Mineno‡and Tadanori Mizuno*

†Graduate School of Informatics, Shizuoka University, Japan
‡Faculty of Informatics, Shizuoka University, Japan
* Graduate School of Science and Technology, Shizuoka University, Japan
{aoi, mineno, mizuno}@mizulab.net

Abstract - Reprogramming sensor nodes is important for managing sensor networks. Since the number of deployments is increasing, the need for an efficient and reliable reprogramming service is growing. In this paper, we outline a framework and analyze several reprogramming challenges. We then describe a comprehensive survey of the state-of-the-art reprogramming systems, and discuss the impact of several design factors.

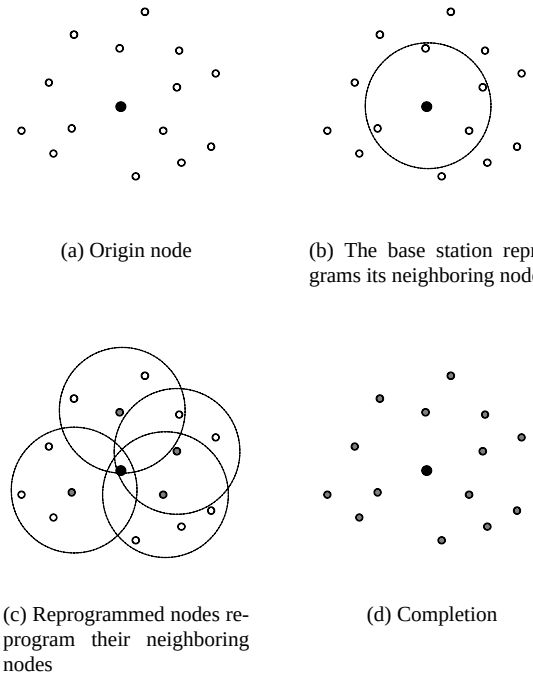
Keywords: Wireless sensor networks, Wireless reprogramming, Software/firmware update, Data dissemination, Encoding/decoding

1 Introduction

Advances in MEMS and low power wireless communication technology have led to the development of wireless sensor networks (WSN). A typical WSN consists of a number of small battery-powered sensor nodes that sense, collect, and transfer various data autonomously. There are many WSN applications and services, including structural monitoring, security, and position tracking. These networks include state-of-the-art technologies (ad-hoc network routing, data processing, position estimation, etc.), and these technologies are implemented as specific code on the sensor nodes. These technologies are highly advanced and still developing. Therefore, these codes will be modified or extended in the future for long-running applications using WSNs. Thus, a method to efficiently reprogram many deployed sensor nodes is necessary.

Wireless reprogramming has been extensively researched [1]. Figure 1 illustrates a process of wireless reprogramming. To reprogram WSN new code is distributed to many sensor nodes using wireless multihop communication. The purpose of general protocols in WSNs is data acquisition: the base station aggregates lots of small amounts of data from the edge nodes. In contrast, the purpose of wireless reprogramming protocols is data dissemination: the base station disseminates large data packets to the entire network. Data dissemination is usually used by system administrators for reprogramming programs on sensor nodes.

Reprogramming protocols now mainly focus on software data dissemination. But it is also necessary to discuss about reprogramming phase. There are several challenges to make an update file and reprogram codes on sensor nodes. During reprogramming, the performance of the WSN is greatly degraded, so reprogramming time should be shortened as much as possible. Also as the resources of the sensor nodes are constrained, energy consumption during reprogramming should be minimized.



(a) Origin node

(b) The base station reprograms its neighboring nodes

(c) Reprogrammed nodes reprogram their neighboring nodes

(d) Completion

Figure 1: Wireless reprogramming

This paper provides the reprogramming framework, design challenges, existing systems and approaches.

2 Wireless Reprogramming Framework

There are several major functions related to wireless reprogramming. Figure 2 shows a reprogramming framework.

Management interface: The interface is installed at a base station for system administrators to manage reprogramming tasks. Administrators can initiate/monitor/cancel reprogramming, set various parameters, and so on.

Version control: A version cotrol database manages various program versions that include past programs and current running programs on sensor nodes.

Update type: If an entire update is required, new code is encoded without modification. If administrators require an incremental update, a delta file is made and then a delta is encoded.

Encoding/decoding: Encoding/decoding is used to reduce

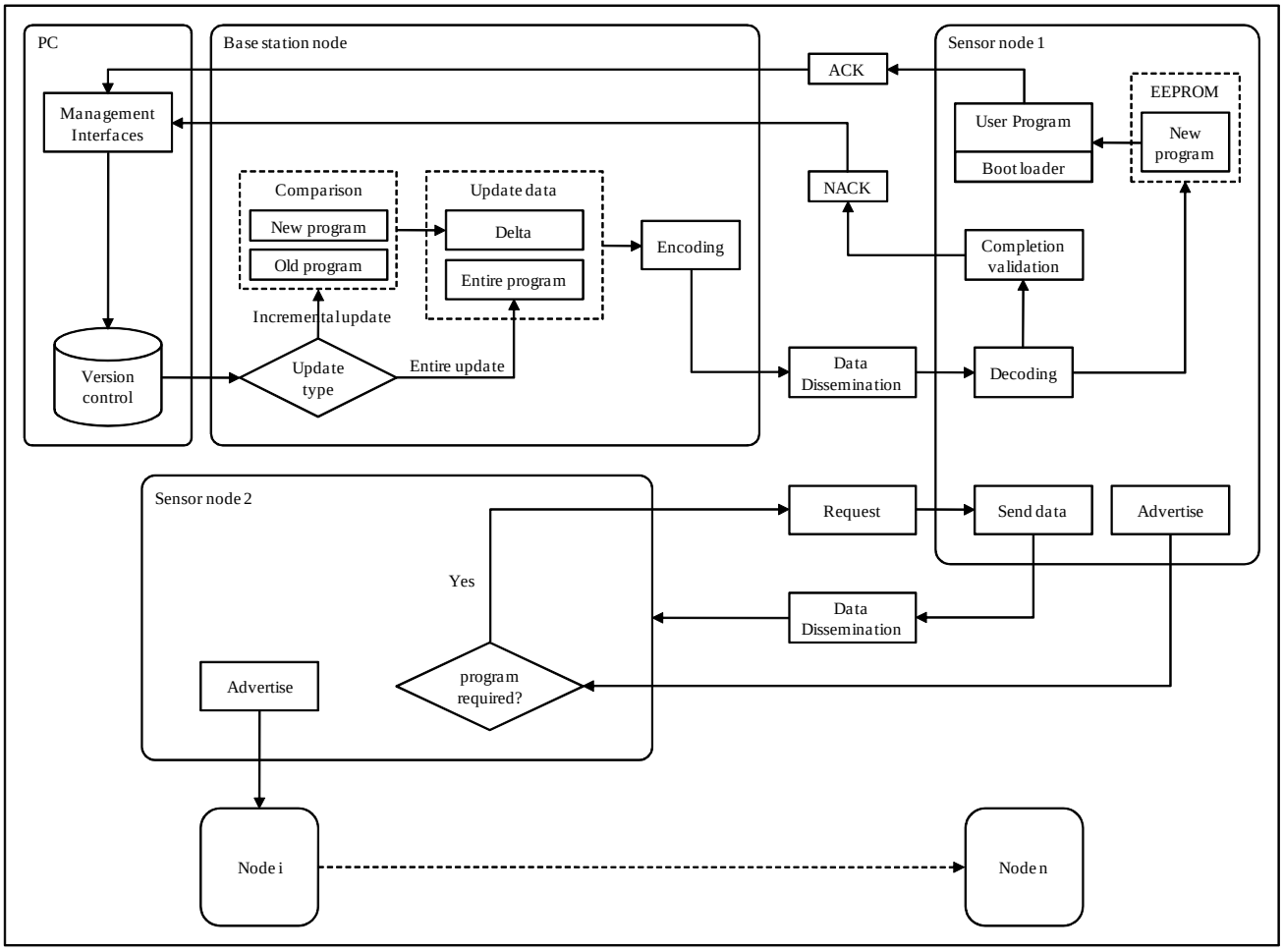


Figure 2: A wireless reprogramming framework

the size of update data. To shorten completion time or minimize energy consumption, small transferred data is desiable. A sender encodes a new program image into data packets and sends it. A receiver decodes these packets and rebuild the image.

Data dissemination: Reprogramming protocols disseminates a program image from base stations to targeted nodes. This function is very important to achieve design goals in reprogramming described in follow sections.

Completion validation: The new program data should be delivered withour errors. Completion validation assure a correct reprogramming.

ACK/NACK: Acknowledgment or negative acknowledgment is send from sensor nodes to administrators. System administrators can get the condition of reprogramming on each sensor node.

Switching: Sensor nodes must change the program on themselves with received data. The new program is stored in an external flash memory (EEPROM). Then sensor nodes switch to the new program through a reboot process.

3 Reprogramming Design Challenges

Many wireless reprogramming protocols share design challenges.

Completion time: The reprogramming completion time affects services using WSNs. When we reprogram a network, we have to stop the services and wait until the code update is completed. Thus, we have to minimize the reprogramming completion time.

Energy efficiency: Sensor nodes are usually battery-powered, and the sensor node battery provides the energy used in reprogramming. This battery also supplies energy for computing, communication, and sensing functions. Therefore, reprogramming must be energy-efficient.

Reliability: Reprogramming requires the new code to be delivered throughout the entire network, and the delivered code must be executed correctly on the sensor nodes.

Reprogramming protocols use several techniques to resolve the challenges listed above. In the next section, we deal with several techniques used in state-of-the-art reprogramming protocols to resolve the challenges listed above.

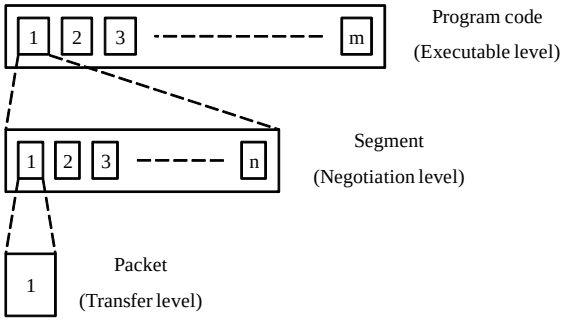


Figure 3: Structure of segment

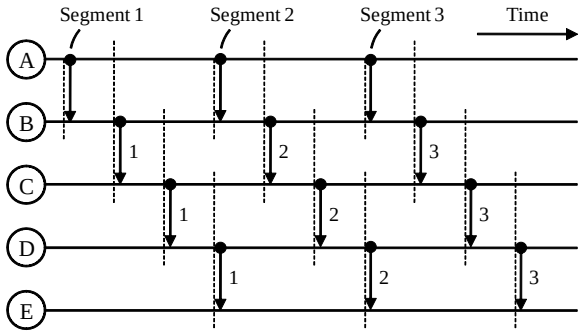


Figure 4: Three pipelining segments in a four-hop network

4 Reprogramming Approaches

4.1 Data dissemination

Pipelining: Pipelining was developed to accelerate reprogramming in multihop networks. [2] [3] Pipelining allows parallel data transfer in networks. In pipelining, a program is divided into several segments as shown in Figure 3, and each segment contains a fixed number of packets. Instead of receiving a whole program, a node becomes a source node after receiving only one segment. Figure 4 shows the process of software distribution in pipelining. There are five sensor nodes deployed linearly in Figure 4. A dashed line represents the interference range, and a solid arrow represents the reliable communication range. To avoid the hidden terminal problem, the parallel data should be transferred with at least three hops spacing. In Figure 4, while D is sending segment 1 to E, A is sending segment 2 to B simultaneously. Thus, pipelining can transfer program codes fast by overlapping the segments. Particularly for transmitting a large program, pipelining can reduce the completion time significantly. However, pipelining has detrimental characteristics. Pipelining works well when there is a large number of hops between base stations and farthest destination nodes. In contrast, having a small number of hops causes delay in code distribution [4].

Negotiation: Negotiation is used to avoid data redundancy and improve reprogramming reliability. As explained above, pipelining is done through segmentation. After the data is segmented, it is necessary to avoid broadcast storms that are caused by dealing with a large number of segments. A negotiation scheme was developed in SPIN [5]. This scheme uses three-way handshakes between senders and receivers. Figure 5 shows the three-stage handshake. There are three types of messages (ADV, REQ, and DATA) in simple negotiation protocols. First, the source node (A) sends an ADV message, which includes its received segment information, to neighboring nodes (B). Second, if the destination node receives the ADV message, it compares its own segment with the received segment information and decides whether it needs the segment sent by the source node. If the segment is needed, the receiver requests the segment from the source node by sending an REQ message. Finally, if the source node receives the REQ message from the destination node, it forwards a requested DATA message. In this scheme, the source node knows which segment is requested before sending it out. This reduces data redundancy. Although the negotiation scheme brings advantages for pipelining, there are some detriments that are caused by control packets. A three-way handshake needs at least two control packets (ADV and REQ messages) to send DATA messages. Thus, if we increase the number of segments to accelerate code distribution, control packets increase at the same time. This affects the energy efficiency and reliability. First, sending messages is one of the most energy-consuming actions for sensor nodes. Second, when a lot of messages are sent, message collisions may occur. For these reason, it is necessary to adjust the number of divided segments.

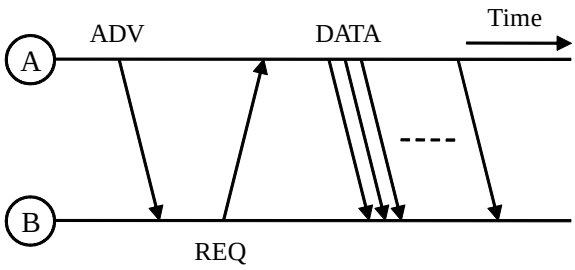
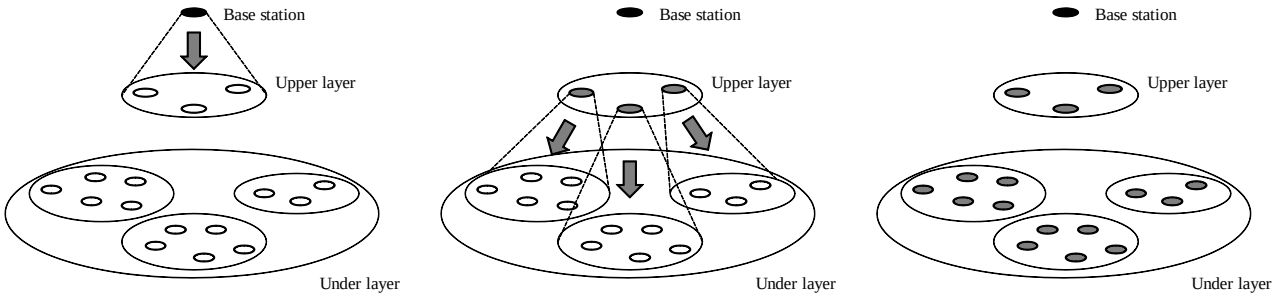


Figure 5: Three-stage handshake

Hierarchical reprogramming: Hierarchical reprogramming accelerates software distribution and reduces the number of control packets. Firecracker [6] and Sprinkler [7] are known as hierarchical reprogramming protocols. Figure 6 illustrates their operation. First, the base station sends program codes to nodes in the upper layer of the node hierarchy (i.e., pseudo-base stations). Then the pseudo-base stations distribute codes to other nodes in their local areas. Except for Firecracker and Sprinkler, most reprogramming protocols start distributing software from a single base station in the network and assume no hierarchy. Hierarchical reprogramming protocols improve reprogramming efficiency, but the decision method for the placement of pseudo-base stations has not been discussed. Deploying base stations in suitable places greatly improves the efficiency of the reprogramming.

Facility location problem: Constructing an efficient network is one of the challenges for sensor networks. Although there are several definitions of efficiency, we consider an efficient network to be a network that can accelerate software dis-



(a) Reprogram upper layer (b) Reprogram under layer (c) Completion

Figure 6: Example of hierarchical reprogramming protocol behavior

tribution, reduce network traffic, or reduce power consumption of sensor nodes. Network construction refers to the placement of base stations. The placement of base stations is treated as a facility location problem. A facility location problem is a mathematical problem about deciding the most suitable location for facilities (e.g., educational facilities, medical centers, department stores, etc.) in terms of economic efficiency or convenience. These problems can be treated as packing problems, covering problems, or problems involving Voronoi diagrams.

4.2 Encoding/decoding

Completion time and energy efficiency of reprogramming is depend on the amount of distributed data. To minimize the size of update data, it is necessary to consider about a method of encoding and decoding.

Incremental reprogramming: In practice, consider about incremental changes to several functions, or modification of the parameters that control current functionality, there is a little difference between the current running program and new program. Since difference is much smaller than the entire program, incremental reporgramming is reasonable way to update the sensor nodes. The purpose of incremental reprogramming is to transfer a small delta (difference between the old and the new program) so that shorten completion time and reduce energy consumption.

Dynamic linking: The design of incremental reprogramming have several challenges. TinyOS [8] that a component-based operating system and platform is widely used. This platform does not support dynamic linking of software components on a node. Thus this system requires transferring entire program data for reprogramming. Other operating systems that represented by SOS [9] and Contiki [10] support dynamic linking.

Virtual machine: Systems such as Maté [11] and ASVM [12] propose virtual machines which run on sensor nodes. These systems provide efficient reprogramming, since the virtual machine code is more compact than the native code. However, they trade off, less efficient execution than native code.

Rsync: Rsync [13] algorithm was originally developed to update binary data between computers over a low bandwidth network. It divides the binary data into fixed size blocks, and then both sender and receiver compute the pair (Checksum, MD4) over each block. However, generally, sensor nodes cannot compute MD4 because its expensiveness. Therefore modified rsync algorithms have been developed to execute on sensor nodes [14] [15] [16].

5 Reprogramming System Design

Here we overview the reprogramming protocol designs. Several reprogramming protocols have been discussed and designed in the last few years, as summarized in Table 1.

Single-hop or multihop: The earliest reprogramming protocols, such as XNP [17], distribute program codes only within the radio communication range of a base station. Recently, many reprogramming protocols target at multihop scenarios. Multihop code dissemination protocols are epidemic, and need to address important issues such as efficiency and scalability.

MAC layer: Most reprogramming protocols use CSMA MAC. Sprinkler [7] and CORD [22] uses TDMA MAC to reduce contention and achieve higher throughput. However, TDMA demands clock synchronization of whole network, and its implementation is much more difficult than CSMA. Therefore TDMA is not currently available for most sensor platforms.

Node hierarchy: Firecracker [6], Sprinkler and CORD support hierarchical reprogramming.

Scope selection: Most reprogramming protocols have no scope selection scheme and disseminate one program to whole network. Aqueduct [20] supports scope selection for targeted nodes. When this protocol reprogram targeted nodes, the number of routing nodes is minimized and the wasted energy is reduced.

Pipelining: Deluge [2], MNP [3], Zephyr [15], Hermes [16], Aqueduct, ReMo [21] and CORD support pipelining to accelerate code distribution in multihop networks.

Sleep scheme: MNP and CORD support a sleep scheme. In these protocols, nodes that are not involved in receiving or transmitting program codes can turn off their radios to reduce their energy consumption.

Mobile environment: Most reprogramming systems are not suited to mobile sensor networks. ReMo supports mobile

Table 1: Comparison of reprogramming protocols

Name	Hop	MAC	Hierarchy	Scope	Pipelining	Sleep	Mobile	Encoding/decoding
Deluge [2]	Multihop	CSMA	No	Whole network	Yes	No	No	Entire program
MNP [3]	Multihop	CSMA	No	Whole network	Yes	Yes	No	Entire program
Firecracker [6]	Multihop	CSMA	Yes	Whole network	No	No	No	Entire program
Sprinkler [7]	Multihop	TDMA	Yes	Whole network	No	No	No	Entire program
Incremental [14]	Single-hop	CSMA	No	Whole network	No	No	No	Platdorm-independent patch
Zephyr [15]	Multihop	CSMA	No	Whole network	Yes	No	No	Delta script
Hermes [16]	Multihop	CSMA	No	Whole network	Yes	No	No	Delta script
XNP [17]	Single-hop	CSMA	No	Whole network	No	No	No	Entire program
MOAP [18]	Multihop	CSMA	No	Whole network	No	No	No	Entire program
Trickle [19]	Multihop	CSMA	No	Whole network	No	No	No	Maté script
Aqueduct [20]	Multihop	CSMA	No	Selected nodes	Yes	No	No	Entire program
ReMo [21]	Multihop	CSMA	No	Whole network	Yes	No	Yes	Entire program
CORD [22]	Multihop	TDMA	Yes	Whole network	Yes	Yes	No	Entire program

environment without making any assumptions on the location of nodes. This protocol estimates link qualities and relative distances with neighboring nodes in order to select the best node for code exchange. Additionally, ReMo downloads segments regardless of their order, thus, exploiting the nodes mobility and facilitating a fast transfer of the program code.

Encoding/decoding: Most reprogramming systems transfer the full program image through the radio communication. On current hardware platforms, radio subsystem is one of the costly functionality in respect of energy consumption. Moreover, a program image is much larger than data that sensor nodes normally transmit such as sensed data. Therefore, to reduce disseminated data, encoding/decoding methods have been developed. Incremental reprogramming approaches compare the difference between the current and new program, and transmits only the delta. Incremental network programming [14] uses optimized Rsync algorithm. Zephyr and Hermes proposed byte level comparison between the old and new program. After comparing program, delta script is made and transmitted. Trickle [19] transmits Maté virtual machine scripts instead of native nesC compiled codes since Maté scripts are much smaller and simpler to write. However, Maté scripts are limited to the function of the virtual machine, and are not as flexible as nesC programming.

6 Conclusion

Reprogramming is necessary in long term operation of WSNs to ease the management and maintenance. WSNs are still in the process of development and there is a possibility to evolve rapidly. In the near future, we will see new hardware platforms, new operating systems, and new services or applications. At the same time, reprogramming will be one of important services of WSNs.

In this paper, we presented two prominent approaches for reprogramming, data dissemination and encoding/decoding. Many researches about data dissemination (data routing) have been discussed in recent years, but there are few knowledge about encoding/decoding on a resource-constrained sensor node. Thus, more researches about coding (e.g. network cording, data compression, etc) is required. And then it is desirable to collaborate data dissemination protocols and encoding/decoding methods.

Building adaptive WSN applications through reprogramming is an attractive area. Reprogramming technologies can make WSNs dynamic environment and provide many chances to create various services. After reprogramming becomes widely used, more adaptable and intelligent systems will be presented.

REFERENCES

[1] Q. Wang, Y. Zhu, and L. Cheng: Reprogramming wireless sensor networks: challenges and approaches, IEEE Network, Vol.20, No.3, pp. 48–55 (2006).

[2] J. W. Hui and D. Culler: The Dynamic Behavior of a Data Dissemination Protocol for Network Programming at Scale. Proc. ACM SenSys, pp. 81–94 (2004).

[3] S. S. Kulkarni and L. Wang: MNP: Multihop Network Reprogramming Service for Sensor Networks. Proc. IEEE ICDCS, pp. 7–16 (2005).

[4] A. Hashizume, H. Mineno and T. Mizuno: Base Station Placement for Effective Data Dissemination in Sensor Networks. IPSJ Journal, vol. 51, no. 3, pp. 753–760, (2010).

[5] J. Kulik, W. R. Heinzelman, and H. Balakrishnan: Negotiation-Based Protocols for Disseminating Information in Wireless Sensor Networks. Wireless Networks 8 (2/3), pp. 169–185 (2002).

[6] P. Levis and D. Culler: The Firecracker Protocol. Proc. ACM SIGOPS European Workshop (2004).

[7] V. Naik et al.: Sprinkler: A Reliable and Energy Efficient Data Dissemination Service for Wireless Embedded Devices. Proc. IEEE RTSS, pp. 277–286 (2005).

[8] TinyOS Community Forum: http://www.tinyos.net/

[9] C. Han, R. Rengaswamy, R. Shea, E. Kohler, and M. Srivastava: SOS: A dynamic operating system for sensor networks. Mobisys, pp. 163–176 (2005).

[10] A. Dunkels, B. Gronvall, and T. Voigt: Contiki-a lightweight and flexible operating system for tiny networked sensors. IEEE Emnets, pp. 455–462 (2004).

[11] P. Levis and D. Culler: Maté: a tiny virtual machine for sensor networks. ACM SIGOPS Operating Systems Review, pp. 85–95 (2002).

[12] P. Levis, D. Gay, and D. Culler: Active sensor networks. 2nd USENIX/ACM NSDI (2005).

[13] A. Tridgell: Efficient Algorithms for Sorting and Synchronization. PhD thesis, Australian National University (1999).

[14] J. Jeong and D. Culler: Incremental network programming for wireless sensors. IEEE SECON, pp. 25–33 (2004).

[15] R. K. Panta, S. Bagchi, S. P. Midkiff: Zephyr: Efficient Incremental Reprogramming of Sensor Nodes using Function Call Indirections and Difference Computation. USENIX (2009).

[16] R. K. Panta and S. Bagchi: Hermes: Fast and Energy Efficient Incremental Code Updates for Wireless Sensor Networks IEEE INFOCOM, pp. 639–647 (2009).

[17] Crossbow Technology, Inc.: Mote In-Network Programming User Reference. (2003).

[18] T. Stathopoulos, J. Heidemann, and D. Estrin: A remote code update mechanism for wireless sensor networks. CENS, Tech. Rep. (2003).

[19] P. Levis et al.: Trickle: A Self-Regulating Algorithm for Code Propagation and Maintenance in Wireless Sensor Networks. Proc. Networked Systems Design and Implementation, pp. 15–28 (2004).

[20] L. A. Phillips: Aqueduct: Robust and Efficient Code Propagation in Heterogeneous Wireless Sensor Networks. Master's thesis, Univ. CO (2005).

[21] P. De, Y. Liu, and S. K. Das: ReMo : An Energy Efficient Reprogramming Protocol for Mobile Sensor Networks. Proc. IEEE PerCom, pp. 60–69 (2008).

[22] L. Huang and S. Setia: CORD: Energy-efficient Reliable Bulk Data Dissemination in Sensor Networks. Proc. IEEE INFOCOM, pp. 574–582 (2008).

Estimation of User Action and Mobile Phone Situation By Using Active Sensing

Yoh Shiraishi[*], Kouta Sugawara[**], and Osamu Takahashi[*]

[*]School of Systems Information Science, Future University Hakodate, Japan
[**]Graduate School of Systems Information Science, Future University Hakodate, Japan
{siraisi, osamu}@fun.ac.jp

Abstract – We propose an active sensing method that uses a mobile phone's acceleration sensor and its vibration motor as an actuator to estimate the phone's situation and its user's action. This method overcomes the problem of existing methods: despite obtaining accurate action estimation from a single acceleration sensor, they cannot estimate the phone's situation because they utilize differences in acceleration sensor data caused by user actions and estimate the user's action based on that change. Our method vibrates the mobile phone by using its vibration motor under pulse width modulation (PWM) control and extracts the feature quantities from the acceleration sensor data. It then simultaneously estimates the user's action (dynamic items) and phone's situation (static items) with high accuracy from the differences in feature quantities.

Keywords: active sensing, mobile phone, acceleration sensor, context awareness, ubiquitous computing.

1 INTRODUCTION

Recent developments in micro-electromechanical systems (MEMS) technology have led to smaller sensors that are easier to be mounted in a mobile phone. There have been many studies on estimating what a user is doing (user action estimation) by utilizing these sensors, and the research results have been applied to various kinds of context awareness services. Many studies on ubiquitous computing and context awareness have paid attention to the estimation of user actions such as walking and jogging [1-4], and most of these studies used acceleration sensors for the action estimation. Some studies targeted the mobile phone and estimated the user actions by using the phone's acceleration sensor [5-7]. Such estimation items, which are regarded as "dynamic items", are estimated on the basis of the differences in acceleration sensor data caused by user action. However, for context aware services, we must estimate not only such dynamic items but also the situation of the mobile phone. Such estimation items are regarded as "static items" because the acceleration sensor data do not change. In other words, estimating a static item means estimating what object is touching the mobile phone (for examples, table, pocket, hand, bag and so on). Our research aims to estimate both static and dynamic items when a user is not only standing (staying still) but also moving (walking and jogging).

In this paper, we describe an active sensing method for simultaneously estimating the user action (dynamic items) and mobile phone situation (static items). We implemented this method in a prototype system and confirmed its effectiveness. Our method uses the mobile phone's vibration motor under pulse width modulation (PWM) control and the phone's acceleration sensor. It uses the motor to vibrate the phone and acquire the vibration through the acceleration sensor. It estimates the static and dynamic items from the differences in acceleration data.

This paper is organized as follows. Section 2 mentions related work and the requirements of our method. Section 3 describes our method for simultaneously estimating user action and mobile phone situation. Section 4 evaluates this method and presents experimental results. Section 5 concludes by briefly summarizing the main points and mentioning future work.

2 RELATED WORK

Action estimation research can be classified according to two metrics: estimation items (dynamic or static) and sensing method (passive or active sensing). In general, passive sensing is used for estimating dynamic items and active sensing for static items. Whereas existing studies have estimated either the dynamic or static items, our method aims to estimate both dynamic and static items by using active sensing simultaneously (Figure 1).

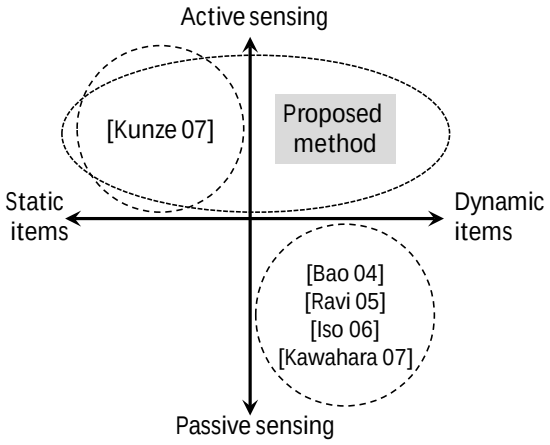


Figure 1: Position of our method.

2.1 Passive Sensing

Passive sensing is a widely used method of sensing for action estimation in existing research [1-7]. Sensors on the user's body and object (e.g., a mobile phone) are moved by the user and by the object to be sensed. Consequently, the acceleration sensor data are changed by the movement. Passive sensing is a method of extracting the feature quantities from the acceleration data and classifying and estimating the types of user action such as standing, walking,

jogging and going up/down stairs. Some of the existing studies used multiple acceleration sensors [1] or a combination of an acceleration sensor and other kinds of sensor [3, 7]. Since mounting many sensors on a user's body will become a burden to the user, it is desirable to use a single small sensor or the mobile phone's sensors for action estimation. There has recently been an increase in the amount of research on action estimation using a single 3-axis acceleration sensor on the user's body or the mobile terminal [4-6].

2.2 Active Sensing

Active sensing acts on the environment surrounding the human and objects through vibration, sound, and light. As a result, it detects differences in vibration, reverberating echo, and illumination intensity with sensors. Active sensing is a method of estimating the situation of the environment from the differences. It does not use acceleration sensor data at the time when the sensors on the user's body and object are moved by the user and object. Existing passive sensing methods have trouble estimating static items such as the situation of a mobile phone (on the table, in a bag, or in a hand) because the sensor does not move and there are no differences in the sensing data over time. However, they can estimate static items by considering differences in the environment caused by, for example, vibration generated by active sensing.

Figure 2 shows changes in the resultant value of the obtained sensor data over time when it executes passive and active sensing for a mobile phone on a table and obtains the phone's vibration with by using the phone's acceleration sensor. The horizontal axis shows data number obtained from the acceleration sensor and the number means the elapsed time. In this figure, we can find changes in acceleration data of active sensing because of the vibration caused by the actuation though there are few changes in acceleration data of passive sensing.

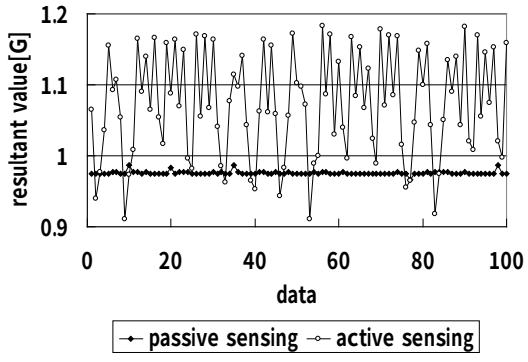


Figure 2: Changes in acceleration sensor data when passive and active sensing is executed.

Kunze [8] used a mobile phone with an acceleration sensor, vibration motor, and microphone and proposed a method for estimating the phone's location by using the vibration and the beep sound.

2.3 Requirements and Problems

Simultaneous estimation of static and dynamic items has some requirements.

- Requirement 1: reduce the data collection time
- Requirement 2: consider the actuator's environmental limitations
- Requirement 3: consider the vibration motor's performance limitations
- Requirement 4: increase the number of feature quantities for better estimation

To reduce the collection time (requirement 1), we assumed instantaneous situations such as the ringing of a mobile phone, for which sample data collection that takes several tens of seconds is inadequate. Estimation using instantaneous data can reduce the time for generating one supervised datum and the time for collecting all necessary supervised data. This has the advantage of reducing the user load for generating supervised data. Consequently, the time taken to collect sample data when a mobile phone is ringing can be reduced to within the range that enables estimation; for this, sample data must be collected with a few seconds. Our method uses sample data collected in 2.56 seconds for estimation as mentioned later.

To consider the actuator's environmental limitations (requirement 2), Kunze [8] used a vibration motor and microphone as actuators for active sensing and the ring beep sound. However, there are many actual situations such as theaters, meeting rooms and public vehicles where mobile phone ring tones should not be used, so use of the beep sound is undesirable. On the other hand, the vibration motor in a mobile phone is often used as a silent alert in the phone's manner mode because the vibration noise level is low. Consequently, we use only the vibration motor and acceleration sensor to achieve highly accurate estimation.

Consideration of the motor's performance limitations (requirement 3) involves a tradeoff between size and performance. The sampling rate of an acceleration sensor small enough to be mounted in a mobile phone is about 200 Hz (Table 1). However, vibration motors that can be mounted in a mobile phone currently run at 9500–13,000 rpm, which corresponds to a motor frequency of 150–200 Hz using the conversion factor of one rotation to one cycle (Table 2).

Table 1: Specifications of acceleration sensor.

Product name	WAA-001
Manufacturer	Wireless technologies, Inc.
Axis	3-axis
Maximum acceleration	±3 [g]
Communication standard	IEEE 802.15.1 (Bluetooth)
Maximum sampling rate	200 Hz

Table 2: Specifications of vibration motors.

Manufacturer	Rated speed	
	rpm	Frequency [Hz]
SHICOH, Co., Ltd.	10,000 ±1800	166 ±30
Minebea Co., Ltd.	11,000	183
SANYO Seimitsu	9500	158

To acquire a precise wave shape, we need a sampling rate that is about twice the vibration frequency according to sampling theorem, which does not let us collect a precise wave form as it is. Consequently, until now there has been a tradeoff between using the beep sound, which is undesirable, and using the mounted acceleration sensor, which cannot acquire the vibration with sufficient precision. We solved this tradeoff problem about environmental and performance limitations of the vibration motor and acceleration sensor, as described later.

To increase the number of feature quantities for better estimation (requirement 4), we considered the feature quantities sufficient for high-accuracy estimation under limitations such those of requirements 1–3.

3 PROPOSED METHOD

3.1 System Overview

Our method uses a vibration motor under PWM control for obtaining accurate vibration data. The differences in acceleration data caused by the vibration motor are mainly used for estimating static items. An object such as a table, pocket, or hand touching a mobile phone vibrates because of the phone's vibration motor, and this vibration propagates back to the phone and the phone's acceleration sensor measures the vibration data. The acceleration data are processed by fast Fourier transform (FFT) and the feature quantities are extracted from the FFT result. Our system estimates both dynamic and static items on the basis of clustering using the feature quantities. An overview of our system is shown in Figure 3.

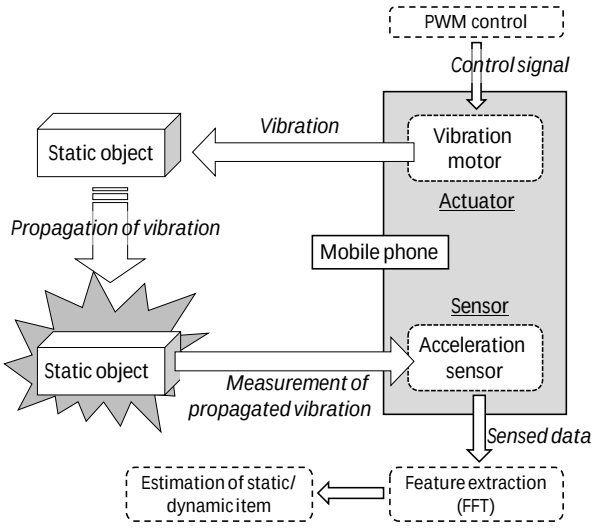


Figure 3: System overview.

3.2 PWM Control

The frequency (number of rotations) of vibration motors commonly used for mobile phones exceeds the sampling rate of the acceleration sensors that can be mounted in mobile phones. Consequently, these sensors cannot acquire the precise wave shape. To solve this problem, we reduce the vibration (namely, the number of rotations). We use PWM control as a rotation control mechanism. PWM control is a motor control technique, and its circuit design and implementation are relatively easy.

The PWM control operates with a duty ratio that is the ratio of the pulse wave's on and off power and adjusts the amplitude of the motor's applied voltage. As a result, it can make the number of motor rotations either high or low. The duty ratio is defined as

$$D = \frac{t}{T}, \qquad (1)$$

where D is the duty ratio, t is the pulse width, and T is the cycle number.

When D is 0.25, the voltage applied to the motor decreases to 1/4 of the original. As shown in Figure 4, the actual applied voltage decreases and the number of motor rotations is reduced.

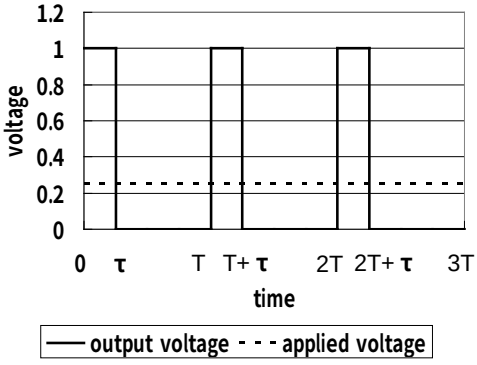


Figure 4: Example of PWM (D=0.25).

By adjusting the duty ratio to suit the number of rotations that can be sampled, we can acquire the precise wave shape at the sampling rate of existing acceleration sensors. However, just after staring operations, PWM control cannot obtain a sufficient number of rotations because the voltage is not stable. Such time lag occurs in PWM control and the lag tends to be longer as the duty ration is smaller. In this paper, we use PWM control with $D = 0.5$ for which the corresponding motor frequency in our experiments is about 108 Hz. To acquire the precise wave shape, we set the sampling rate to the maximum value (200 Hz) of the acceleration sensor on the basis of sampling theorem.

3.3 Estimation Items

Our method estimates basic user actions as dynamic items. We consider three basic actions: standing, walking, jogging. As static items, it estimates the phone's situation (What object is touching the phone? How is the user keeping the phone?). We consider five representative situations: a mobile phone in a pocket (Pocket), in a hand (Hand), on a table (Table), in a bag on the floor (Bag-put), and in a bag in hand (Bag-carry). When the user is walking or jogging, he or she does not put a bag containing the phone on the floor and the mobile phone is not on a table either. Consequently, our method estimates 11 different combinations of dynamic

items (actions) and static items (objects), as shown in Table 3.

Table 3: Dynamic and static estimation items.

Estimation items (action-object)	
Stand-Pocket	Walking-Pocket
Stand-Hand	Walking-Hand
Stand-Table	Walking-Bag-carry
Stand-Bag-put	Jogging-Pocket
Stand-Bag-carry	Jogging-Hand
	Jogging-Bag-carry

3.4 Feature Quantities for Estimation

We use FFT as a frequency analysis method for extracting feature quantities. According to existing research [2], the frequency range that is effective to estimate dynamic items such as walking and jogging is about 1.5-3.5 Hz. When analyzing a wave shape for walking that has slow movement and small frequency, it is better to collect sample data for as long a time as possible for obtaining a wave shape with periodicity. Figure 5 shows changes (namely, wave shape) in the resultant value of acceleration data when a user is walking and a mobile phone is in a pocket. The sampling rate is 200 Hz.

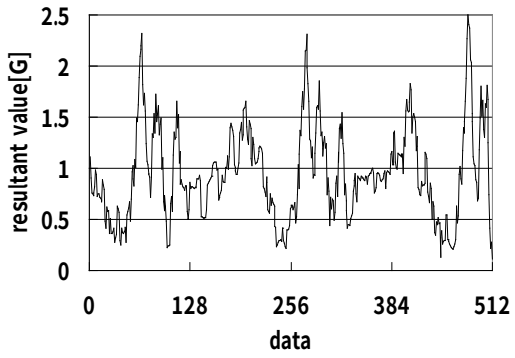


Figure 5: Example of wave shape of acceleration data.

In order to obtain 256 and 512 sample data, it takes 1.28 and 2.56 seconds respectively when the sampling rate is 200 Hz. We can find only one periodical wave shape in 1.28 seconds (256 sample data), but two periodical shapes in 2.56 seconds (512 sample data). Therefore, we collect and use 512 sample data for frequency analysis and estimation.

Our method uses the following feature quantities: values of the integral of the power spectrum (PS) of the FFT for each 10-Hz range from 0 to 100 Hz, maximum values of PS for the 0–40-Hz range and 40–100-Hz range, frequency at the maximum value, resultant value of the variance and mobility of the resultant value (Table 4).

In a preliminary experiment, we found that the estimation accuracy was reduced when the range of each PS was narrow because there were few differences in the integral values of the PS among the estimation items. We use the maximum value of PS from 0 to 40 Hz to estimate the dynamic items and from 40 to 100 Hz to estimate the static items. Each feature quantity is effective for either a dynamic or static item. The mobility is the variance of the inner product of data every 40 ms. We use the mobility as well as the resultant value of the variance as measures to indicate the degree of change in acceleration data (namely, the intensity of phone movement).

Table 4: Feature quantities for estimation.

Value of integral of PS	0–10 Hz	10–20 Hz
	20–30 Hz	30–40 Hz
	40–50 Hz	50–60 Hz
	60–70 Hz	70–80 Hz
	80–90 Hz	90–100 Hz
Maximum of PS	From 0 Hz to 40 Hz	From 40 Hz to 100 Hz
Frequency (at maximum of PS)	From 0 Hz to 40 Hz	From 40 Hz to 100 Hz
Variance	Resultant value	Mobility

4 EVALUATION AND DISCUSSION

4.1 Prototype System for Experiments

The prototype terminal for our experiments is shown in Figure 6. An acceleration sensor, vibration motor, and PWM control circuit are implemented in the body of a mobile phone. We used the 3-axis acceleration sensor WAA-001 (Table 2) from Wireless Technologies, Inc. and the vibration motor B2BA (Table 1) from SICOH Co., Ltd. The maximum sampling rate is 200 Hz. By transmitting the PWM control instructions from a personal computer to the terminal via Bluetooth, we could control the terminal's vibration motor. The terminal vibrated under PWM control and the vibration data were measured by the acceleration sensor. The acceleration data were transmitted to the computer, and the computer executed some processes for estimation such as FFT and clustering.

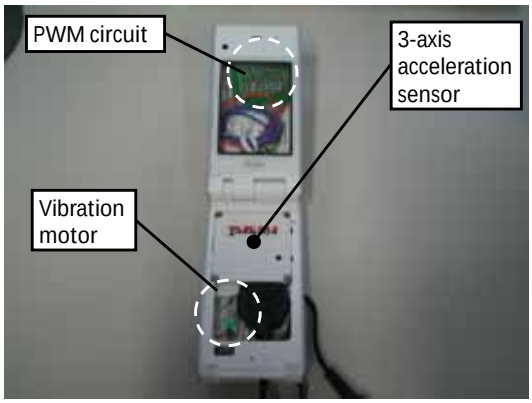


Figure 6: Prototype terminal for our experiments.

4.2 Experiments

Previous studies used various kinds of estimation methods as accuracy estimation methods, such as decision tree, linear discriminant analysis, neural networks, and Bayesian estimation. In this study, we used a data mining tool Weka [9] to calculate the estimation accuracy. We evaluated the

estimation accuracy by using three algorithms implemented on Weka: J48 (based on a decision tree C4.5), three-layer perceptron, and Bayesian network. In this experiment, we simultaneously estimated both static and dynamic items when a user was moving (standing, walking, or jogging).

We collected sample data from ten different combinations except Stand-Table mentioned in Section 3.3 (Table 3). The number of sample data collected for each combination was 100. We prepared tables made of three different materials and collected 50 sample data for each material. The total number of sample data was 1150.

4.3 Evaluation of Significance of Amount of Supervised Data

First, we conducted an experiment to evaluate the significance of the supervised data amount. We used 100 sample data for each estimation item except for Stand-Table. The Stand-Table sample data were reduced from 150 to 100. The total number of evaluation data for this experiment was 1100. We used 110 of these 1100 evaluation data (namely, 10 data for each estimation item) as test data. The remaining 990 were used as supervised data. The supervised data were input and incremented by 110. The estimation accuracy of the results obtained by each algorithm was calculated for each supervised data number. The change in estimation accuracy versus supervised data number for each algorithm is shown in Figure 7.

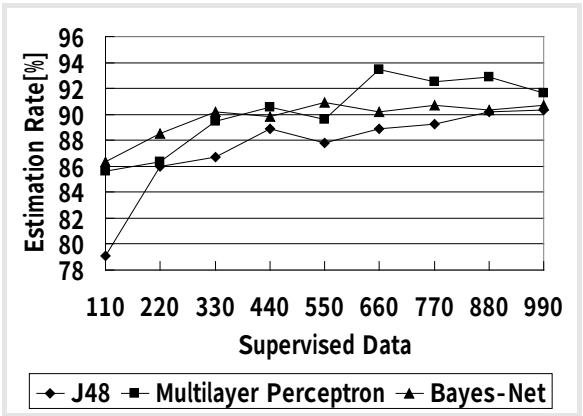


Figure 7: Estimation accuracy versus supervised data number.

The estimation accuracies of all the algorithms tended to increase approximately with increasing supervised data number, and the accuracy was over of 90 %.

In the case of J48, the estimation accuracy was low up to 330 data and gently improved thereafter. We would expect the accuracy to improve for more than 990 supervised data. In the case of a Bayesian network (Bayes-Net), the estimation accuracy was higher than that with J48 up to 990 data. However, changes in the estimation accuracy were very small around 550 data, and the accuracy tended to converge. On the other hand, in the case of the three-layer perceptron (multilayer perceptron), the estimation accuracy gently increased from 110 data and tended to decrease from 660 data. It is likely that using more than 990 data had a negative effect on estimation improvement.

Consequently, in the case of a system constructed for an individual user without supervised data when the system starts operating, it is advisable to use a Bayesian network in the early phase of operations that the system has fewer supervised data and to switch to J48 when the system has enough data. The estimation accuracy of the three-layer perceptron was higher than that of the other two algorithms from 660 data. However, the accuracy might have been reduced by overfitting, as mentioned above, so consideration of the settings such as feature quantity adjustment is required.

4.4 Evaluation of Estimation Accuracy

Next, we conducted an experiment to evaluate our method's estimation accuracy. In this experiment, we used 1150 supervised data as evaluation data. We chose to use the 10-fold cross-validation method [10], which can make the precision error small because the error can easily become large depending on the test data if there are not enough data.

Table 5 and 6 show the estimation accuracy of 11 different estimation items and the average accuracy when 16 feature quantities in Table 4 were used. The average estimation accuracy for J46 was 91.32 %, which is higher than that for other algorithms. However, the difference in accuracy between J48 and the others was under 0.4 %. The accuracies for Walking-Hand, Walking-Bag-carry, Stand-Hand, and Stand-Bag-put were about 10–20 % lower than those of other estimation items for every algorithm.

Table 5: Estimation accuracy (%) (walking, jogging).

Class / Learning algorithm	Walking			Jogging		
	Pocket	Hand	Bag-carry	Pocket	Hand	Bag-carry
J48	96.15	81.19	82.45	99.40	99.00	97.72
Multilayer perceptron	97.95	91.49	88.18	99.85	96.99	94.96
Bayes-Net	98.40	86.47	87.44	99.90	96.65	96.89

Table 6: Estimation accuracy (%) (standing, total average).

Class / Learning algorithm	Stand					Total average
	Pocket	Hand	Table	Bag-put	Bag-carry	
J48	92.23	85.98	92.95	83.95	92.63	91.32
Multilayer perceptron	89.64	83.48	91.20	76.19	93.62	91.23
Bayes-Net	90.70	77.72	92.84	76.50	94.74	90.99

Table 7: Breakdown of estimation results.

	Number of estimated item												
		a	b	c	d	e	f	g	h	i	j	k	
Targeted estimation item	a	137	4		7	2							a: Stand-Table
	b	5	91		3	1							b: Stand-Pocket
	c			93	2	5							c: Stand-Bag-carry
	d	6	3	3	85	3							d: Stand-Bag-put
	e	1	1	3	5	89		1					e: Stand-Hand
	f						96		2	1	1		f: Walking-Pocket
	g						1	80	19				g: Walking-Bag-carry
	h						1	18	81				h: Walking-Hand
	i									100			i: Jogging-Pocket
	j										100		j: Jogging-Bag-carry
	k										2	98	k: Jogging-Hand

A breakdown of the estimation results for J48 is given in Table 7. Each cell shows the number of that estimated item (estimated result) for each targeted estimation item. Cells on the diagonal show the number of correctly estimated items, while off-diagonal cells show false estimates There were many false estimations for Walking-Hand and Walking-Bag-carry, and the results show that the feature quantities were insufficient to estimate these static items when a user is walking. It can be said that estimating Walking-Hand and Walking-Bag-carry is difficult.

The estimation results for only dynamic items are shown in Table 8. The ones for J48 are almost correct. It can be said that it is difficult to estimate static items under the effects of user actions (dynamic items).

Table 8: Estimation results for dynamic items (%).

	Stand	Walking	Jogging	Total avg.
J48	99.90	99.20	99.04	99.49
Multilayer perceptron	100.00	96.94	96.93	98.40
Bayes-Net	100.00	97.40	97.27	98.60

The estimation results for only dynamic items are shown in Table 8. The ones for J48 are almost correct. It can be said that it is difficult to estimate static items under the effects of user actions (dynamic items).

4.5 Discussion

We investigated which feature quantities were used in the case of estimation by J48. Figure 8 show the structure of a decision tree after 10 trials in this case. Each level of the hierarchy of this figure means the depth and branch of the decision tree. 0-10 Hz, 10-20 Hz, 20-30 Hz, 60-70 Hz means the value of integral of PS for each frequency range respectively. LowMaxHz and HiMaxHz are the frequency at maximum value of PS for 0-40 Hz and 40-10 Hz range.

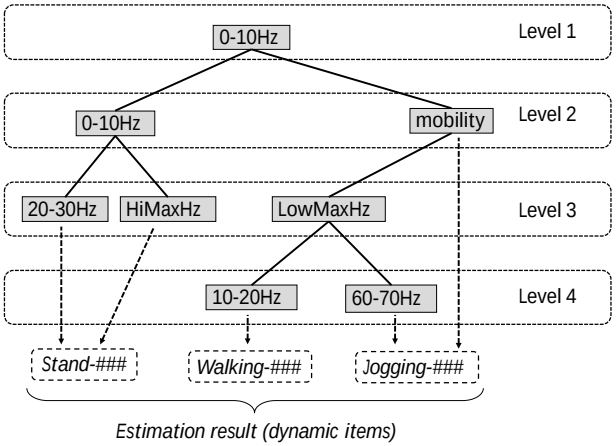


Figure 8: The structure of a decision tree after 10 trials.

It used the value of integral of PS for 0-10 Hz range as the first branch of the tree and it can classify the estimation results to standing or walking/jogging at this branch. Also, the estimated items were classified to walking or jogging at the branch (LowMaxHz) of level 3. Consequently, it can be said that there is little misestimation among dynamic estimation items (standing, walking, jogging). Whereas dynamic items are classified and estimated at the branches from level 1 to level 4, static items are classified at the lower level. We consider that the feature quantities used at the levels from 1 to 4 are effective for estimating dynamic items. On the other hand, many kinds of feature quantities were used at the lower level for estimating static items.

We also confirmed that three feature quantities (values of the integral of PS for the 30-40Hz and 80-90 Hz range and the variance of the resultant value) were not used in this investigation. Consequently, it can be said that these feature quantities were not effective in J48. Tables 9 and 10 show the estimation accuracy when 13 feature quantities except the above three quantities were used.

Table 9: Estimation accuracy (%) when using 13 feature quantities (walking, jogging).

Class / Learning algorithm	Walking			Jogging		
	Pocket	Hand	Bag-carry	Pocket	Hand	Bag-carry
J48	96.53	81.89	83.97	99.40	99.00	97.72
Multilayer perceptron	97.06	88.44	86.77	99.75	94.88	93.29
Bayes-Net	98.50	87.06	87.31	100.00	96.82	96.61

Table 10: Estimation accuracy (%) when using 13 feature quantities (standing, total average).

Class / Learning algorithm	Stand					Total average
	Pocket	Hand	Table	Bag-put	Bag-carry	
J48	92.59	85.66	93	84.3	93.59	91.59
Multilayer perceptron	90.72	81.75	90.06	77.61	93.02	90.41
Bayes-Net	91.41	80.44	93.68	78.52	95.45	91.65

The estimation accuracy of J48 and Bayesian network was slightly improved than that using 16 feature quantities. On the other hand, the accuracy of multilayer perceptron was reduced. Eliminating unnecessary feature quantities in the case of using a decision tree such as J48 is easy and effective. However, it is difficult to judge which feature quantities is effective in the case using multilayer perceptron. Selecting multiple perceptron as a learning algorithm of our method is not practical.

5 CONCLUSION AND FUTURE WORK

In this paper, we proposed a method of simultaneously estimating dynamic items (user actions) and static items (mobile phone's situation) by active sensing with a vibration motor and PWM control. The experimental results gave an estimation accuracy of 91.32 % for J48 and 90.99 % for a Bayesian network. Since a mobile phone is moved by the user's actions, estimating not only the mobile phone situation but also the user action is required for context-aware services. The number of the estimation items increases because of the addition of dynamic items as estimation items. Under such difficult conditions, which make estimation difficult, we were able to estimate static and dynamic items with high accuracy. However, the estimation accuracy of some estimation items such as Walking-Bag-carry and Stand-Hand was only about 70–80 %, so there is room for improvement.

One topic for future work is to improve the accuracy of static item estimation. We will also consider increasing the number of both static and dynamic items and discuss an estimation method that does not lose estimation accuracy as the number of estimation items increases. Furthermore, we must consider personalization technologies to improve our method's robustness.

REFERENCES

[1] L. Bao and S. S. Intille, Activity Recognition from User-Annotated Acceleration Data, PERVASIVE 2004, vol. 3001, pp. 1–17 (2004).

[2] N. Kern, S. Antifakos, B. Schiele, and Adrian Schwaninger, A Model for Human Interruptability: Experimental Evaluation and Automatic Estimation from Wearable Sensors, Proceeding of 8th IEEE International Symposium on Wearable Computers (ISQC'04), pp. 158–165 (2005).

[3] J. Lester, T. Choundhury and G. Borriello, A Practical Approach to Recognizing Physical Activities, PERVASIVE 2006, pp.1–16 (2006).

[4] N. Ravi, P. Mysore, M. L. Littman, Activity Recognition from Accelerometer Data, Proceedings of 7th Conference on Innovation Application of Article Intelligence, pp. 1541–1546 (2005).

[5] T. Iso and K. Yamazaki, Gait analyzer based on a cell phone with a single three-axis accelerometer, Proc. 8th Conference on Human-computer interaction with mobile devices and services, pp.141–144 (2006).

[6] Y. Kawahara, H. Kurasawa, H. Morikawa, and T. Aoyama, Recognizing User Context Using Mobile Handsets with Acceleration Sensors, Proc. IEEE International Conference on Portable Information Devices (Portable 2007), pp.1–5 (2007)

[7] D. Siewiorek, A. Smailagic, J. Furukawa, A. Krause, N. Moraveji, K. Reiger and J. Shaffer, SenSay: a context-aware mobile phone, Proc. 7th IEEE International Symposium on Wearable Computers, pp.248–249 (2003).

[8] K. Kunze and P. Lukowicz, Symbolic Object Localization Through Active Sampling of Acceleration and Sound Signatures, UbiComp2007, pp. 163–180 (2007).

[9] Weka Machine Learning Project, Weka: Waikato Environment of Knowledge Analysis, http://www.sc.waikato.ac.nz/~ml/weka/index.html

[10] C.M. Bishop, Pattern recognition and machine learning, Springer (2006).

Development and evaluation of a routing simulator for a Mutually Complementary Network incorporating Wired and Wireless

Naoki Yusa[1], Hiroki Morita[1], Tadanori Mizuno[2], Kunihiro Yamada[1]

[1] Professional Graduate School of Embedded Technology, Tokai University, Japan, [2] Shizuoka University

{9aejp017@mail.tokai-u.ac.jp,yamadaku@tokai.ac.jp}

Abstract **-** Networks, whether wired and wireless, used at homes are now 100% mutually complementary in terms of communication performance, but that performance can be as low as 4.1% if used in a school building, for example. The objective of the study described herein is to improve communication performance using additional routing capability. A simulator was developed to ascertain and evaluate the characteristics of such a mutually complementary network with an added routing capability. Assuming that we have up to 15 s available for practical use of communication, such a judgment was made that up to 11 nodes for every five-area building was about reasonable to use for practical uses of communications. With that fact in mind, a mutually complementary network was shown to be available for use in a school building by allocating every domain unit of 11 nodes to the entire building. This time, by the function was separated into 2, the simulator speed up. For speed-up of the execution time of simulator, decrease the access frequency to HDD, and decrease the process clone of data creating.

Keywords**:** simulation, network, PLC, Zig-Bee, building

1. MUTUALLY COMPLEMENTARY NETWORK COMMUNICATIONS

A mutually complementary network supports communication using two or more different communication methods concurrently to improve the characteristic of the communication performance [1]. Figure 1 presents a diagram showing the concept of that communication method. In the study described herein, Power Line Communication (PLC) [2] was used for wired communication, with Zig-Bee [3] for a wireless communication. For a three-area and 200m^2 ferroconcrete house or condominium, it has been recognized that the use of PLC alone achieves 70% of communication performance, and that the use of Zig-Bee alone gives 82%. The mutual complementary network using those two communication methods concurrently is expected to increase to 94.6% of theoretical communication performance from a mathematical perspective, but it actually attains 100% [1]. As described above, a mutually complementary network, if used for a private homes, is understood to have sufficient communication performance that can be available for practical use. However, its performance for such buildings as schools is only 4.1% in practice, which is far from satisfactory [4]. To confront this problem, a routing method is expected to improve the situation.

A study has already been made of using a routing method to improve communication performance by allowing every node to recognize its location [5]. One problem is the fact that home electronic appliances or office machines that are placed and used at home or in the office are of lower functionality than a PC, for example. To have a network incorporating such equipment, we must therefore use a less-burdened routing method. One idea was to have a simply processed routing capability with no node location information given to each device at all, thus reducing the burden. That routing takes such a method that it tries to communicate with one node; if it finds that it failed, it moves to try every other node one by one in an arbitrary order. For one-time communication, it bypasses the node with which the communication fails. No re-execution would be made in such a case. Because this method is provided with no mutual location information, it is called a location-undefined communication method. Incidentally, the reported idea of using both wired and wireless communications used in that study resembles the mode used for the study described herein [6]. However, an important difference is that it uses a method by which either of the wired or wireless communications is selected for every communication space concerned, which means that it does not always use both communications concurrently. The mutually complementary network described herein has a characteristic of using both wired and wireless communications concurrently.

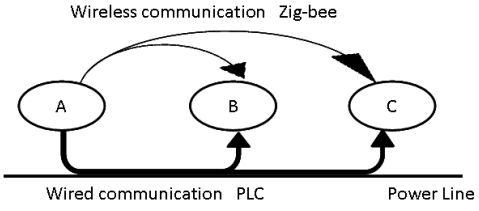


Fig. 1: Concept of a mutually complementary network.

2. NECESSITY FOR SIMULATOR USE

To have a mutually complementary network for a school building, for example, a routing method is taken to improve its communication performance. Here, a location-undefined communication method is used as described above. To accomplish a mutually complementary network with it, the routing performance must be verified first. In other words, we must check to determine whether or not communication can be accomplished within a practically allowable period of time. For that, we must know the averaged communication time period that is taken technically for the shortest and longest routes as well as that for the entire route. However, we understand that vastly numerous combinations of routing exist when we examine the entire route, which indicates to us that we must determine the existing node disposition before anything else. Because it is practically difficult to do that manually, we must proceed to check with the help of a simulator. For the node distribution shown in Figure 2 with nine nodes for a three-area building, the route combinations can be as many as 950 in all, for which we need a simulator for correct modeling.

Each number appearing in Figure 2 represents a serial number given to a node. The electric power supply to Japanese homes is 100V of single-phase three-wire type, separated into phase A and phase B. Communication, if tried using between phase A and the phase B, requires passage through a transformer, which lengthens the communication path distance, thereby increasing the load to its signal transmission, and further tends to pick up more noise. Overall, the communication performance becomes hampered significantly. Furthermore, with the Zig-Bee wireless communication used, we know that its communication performance is degraded by obstacles in its straight communication path line. That might happen often when trying to communicate with the node on another floor of a building, for example. During the simulation we performed, communication performance lower than the threshold value of 50% is taken as 0%, and that higher than the threshold value as 100%. For this reason, communication using PLC is accomplished within the same phase only, in which sense only communication between the node number groups of 1, 3, 5, 7, and 9 and another node groups of 2, 4, 6, and 8 is available. With Zig-Bee used, it is available only between the node number group of 1, 2, 3, and 4 and another group of 5, 6, 7, and 8.

The simulator that is used shows the necessary routing trials made to accomplish communication between the associated nodes. It computes and outputs the minimum number of trials made, the maximum number of trials made, and the averaged value of those trials. The output is selectable either in a CSV format, a text format or an HTML format. The simulator specifications are presented in Table 1.

Table 1: Simulator specification

Specification	
Input data	Number of nodes
	Comm. availability table
Output data	Path for all nodes
	No. of comm. trials for each node path
	Max and min numbers for all comm. trials
	Averaged value of all comm. trials
Output method	CSV format
	Text format
	HTML format

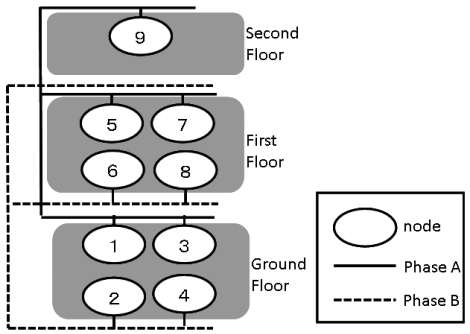


Fig. 2: Example : nine-node distribution in a three-area building.

3. DEVELOPMENT OF A SIMULATOR

The simulator was developed with C language using a currently available standard PC. The total lines of code of the program were 1065; its executable file size was 48 KB. Figure 3 presents the order in which the simulator actions take place. The simulator first detects the state of each evaluated node—where it is located on what floor of the building—and creates a communication availability table based on the phase relation. It then executes the simulator itself with the nodes given. Table 2 shows the communication availability table for nine nodes spread all on the three-area building floors, as shown in Fig. 2.

As described previously, because the communication performance level is determined at the threshold value of 50% with respect to whether the communication would be possible or not, the communication availability table for the node distribution of Fig. 2 is given as that shown in Table 2. In the table, element P represents that the communication is available using PLC, and Z indicates that it is available using Zig-Bee. The element value x indicates that the communication is not available in that position. Although the communication between nodes 1 and 3 of Table. 2 is available using either PLC or Zig-Bee, only the value of P is shown there for reference purposes.

Table 2: Transfer enable table for nine nodes

-	1	2	3	4	5	6	7	8	9
1	-	Z	P	Z	P	x	P	x	P
2	Z	-	Z	P	x	P	x	P	X
3	P	Z	-	Z	P	x	P	x	P
4	Z	P	Z	-	x	P	x	P	X
5	P	X	P	X	-	Z	P	Z	P
6	x	P	x	P	Z	-	Z	P	X
7	P	X	P	X	P	Z	-	Z	P
8	x	P	x	P	Z	P	Z	-	X
9	P	X	P	x	P	x	P	x	-

Figure 4 shows the simulator process sequence. As the simulator process operates, if the communication between two node points is found to be unavailable, then it is bypassed. That process progresses as depicted in Figure 6 as a bypass process of simulator in the way to add a new table for it.

Upon receipt of the information of the nodes and the communication availability table, the simulator creates the Combination List 1 (CL1), which is the list of the nodes telling whether the communication for each pair of them is available or not. First, whether the communication between each pair is directly available or not is checked based on the information of the communication availability table. If any node combination for which no direct communication is available is found, then the simulator halts checking with CL1, creates a new combination list (CL2) to use for a bypass process, and performs a routing process to check whether or not the communication would be available with the route that is newly found. Completing the check on the current combination list, it reverts to the previous list and starts checking it similarly. While repeating such process, it checks all possible combinations in that manner.

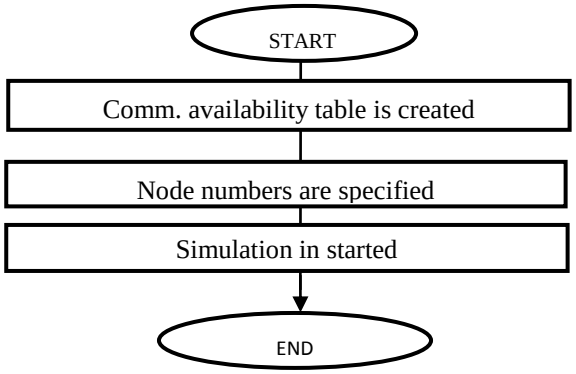


Fig. 3: Sequence of Simulator action

Figure 5 shows working of the simulator. At first, the simulator makes the 1st combination list. Then simulator checks transfer enable table. If can't communication, simulator makes new list name 2nd and make new combination use easy routing algorism. Then the simulator changes it from 1st to 2nd. Then the simulator checks transfer enable table. If the simulator reaches the end of it, change the list above of current list. The simulator continues for finish checking all combination lists.

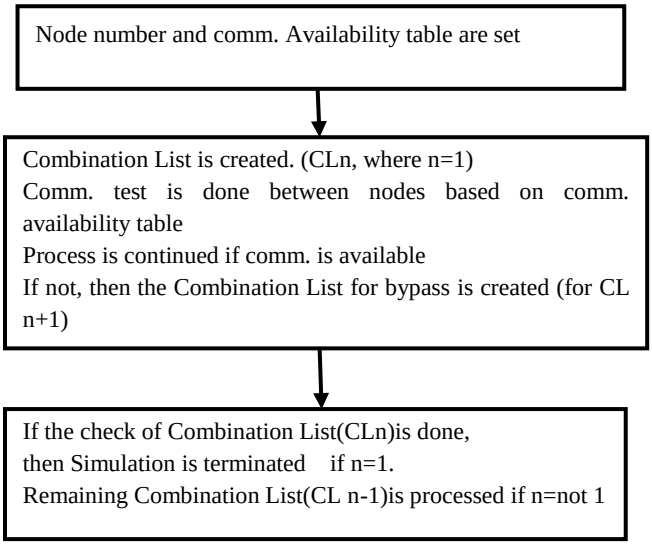


Fig. 4: Simulator process sequence.

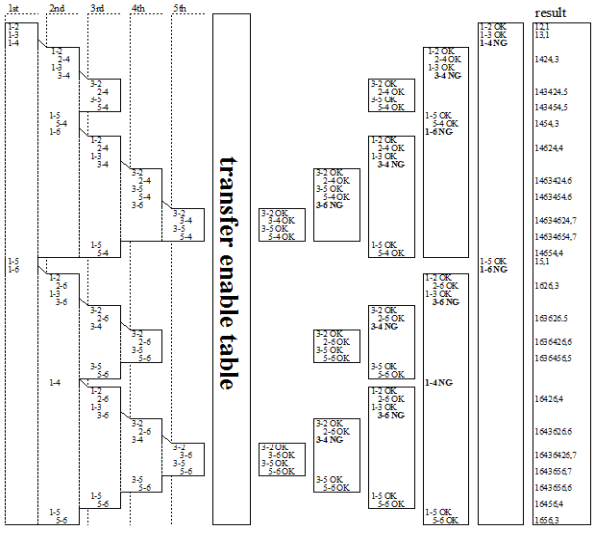


Fig.5: working flow of simulator.

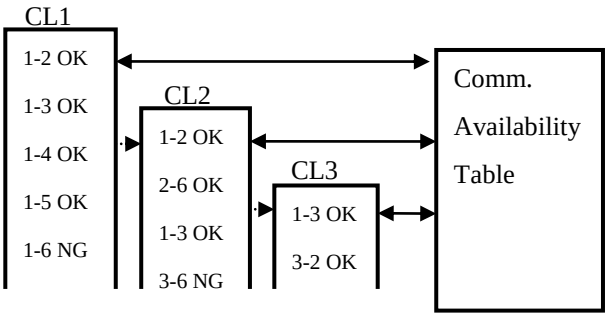


Fig. 6: The bypass process of simulator: CL2 is created because the 1–6 communication of CL1 fails so that it is bypassed. Similarly, it tries to connect via node 3. Because it is still not possible, it further bypasses to move to create CL3

Attempted to shorten the processing time of the simulator. Simulator computes "Number of routes ,passed maximum nodes, and average of passed nodes", and "Passed node number of each route" in number of total nodes n composed by area e. For instance, the processing time of simulator requires 80 hours in case of 11 nodes in 5 areas. For speed-up, "Passed node number of each route" was separated and "Number of routes ,passed maximum nodes, and average of passed nodes" was able to execute in 3 hours. Execution of "Passed node number of each route" needs the data storage to HDD and spends a lot of time(For 77 hours).In addition, I want to try improving.

This simulator spent very long time.it was over than 80 hour in worst case. Its cause was wrote detail of all result to file. This file wrote for comparing with result that human do, and verification simulation process. Optimized program is NOT write detail, and write summary of result (route counts, longest and shortest route length). As a result, the access frequency to HDD was decreased, and it succeeded in shortening time. Fig.7 shows difference of frequency before optimization and after optimization.

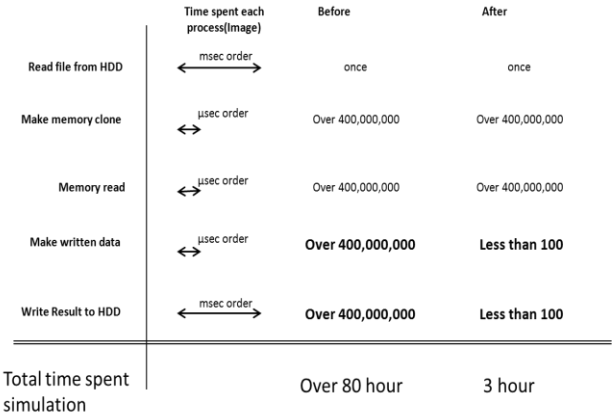


Fig.7: Change of processing count by optimization.

4. SIMURATION RESULTS

Table 3 presents simulation results obtained for eight nodes spread throughout the two-area building. In Table 3, "SN" stands for serial number, each node parenthesized in the path line that is given under "Route" signifies that no data were received there, "adr" stands for the address to, "Tn" stands for the serial number that falls in the same address, "R" for the number of route combinations to come to the same address, "Try" for how many times the communication availability table was referenced, and "TryAv" for the average number of times the communication availability table was referenced in the same address. Figure 8 portrays the relation between the number of nodes and the associated number of routings made based on the simulation results available. Figure 9 shows the total number of routes. Up from nine nodes, node distribution of two kinds was taken in the simulation. The solid line portion represents the case in which nodes are added on the two-area building, and the dotted line portion represents the case in which every new node is gradually added to one floor up. There were 379 route combinations and 7 times of maximum trial for 10 nodes in a two-area building, with 1,697,456 route combinations and 23 maximum trials for 10 nodes in a four-area building. For 11 nodes in a five-area building, there were over 400 million route combinations. The simulation duration lasted over 70 hr. Examining the nodes and the number of routings made in Fig. 6 reveals that up to 29 times routing was made for 11 nodes in a five-area building. Whether the communication is available or not was determined by sending a packet to the node; if no response was returned, then it is judged that the communication was unavailable there. Presuming that it takes 0.5s every time to check whether the communication is available or not, then this example of 29 times of routing takes 14.5 s in all, meaning that up to a 14.5s delay would occur to start communication. Furthermore, for the case of 10 nodes in a four-area building, 11.5s are caused to delay, and for nine nodes in a three-area building, 6s delay. For the 12-node distribution in the two-area building, the maximum number of trials was 13 times, and the time necessary for the routing process was, consequently, 6.5s.

Figure 10 and figure 11 is histogram of simulation result. Horizontal line means number of routing. And vertical line means number of appareling the number of routing. Figure 10 contains 4 nodes for 8 nodes. And figure 11 contains 9 nodes for 11 nodes. 9d for 11d means difference node spreading pattern. The additional nodes set new room without other nodes. So the new nodes can communicate only use PLC. Then the simulator try a lot of pattern to check can communicating nodes. Figure 10 shows less changes because the nodes spread pattern is similar each other. So the result is similar too. However, figure 11 shows big changes. In these cases, node spreads widely than the case of 4 nodes for 8 nodes. So the histogram

shows big differences. This result means the layout of nodes is an important thing to make mutually complementary network using wired and wireless communications. Figure 12 and figure 13 show time spent for complete simulation. In figure 12, Program has not been optimized. In figure 13, Program has been optimized.

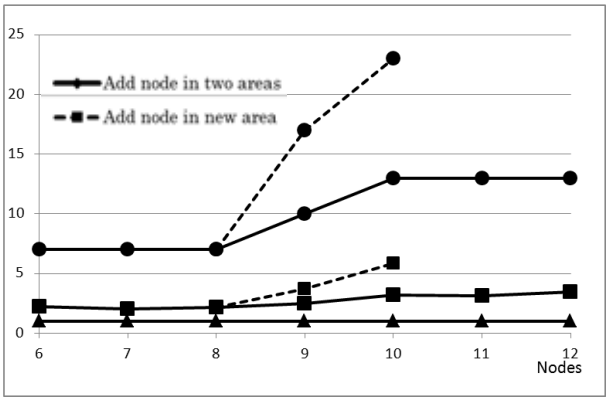


Fig. 8: Node and routing trial relations: ● maximum value, ■ averaged value, and ◆minimum value.

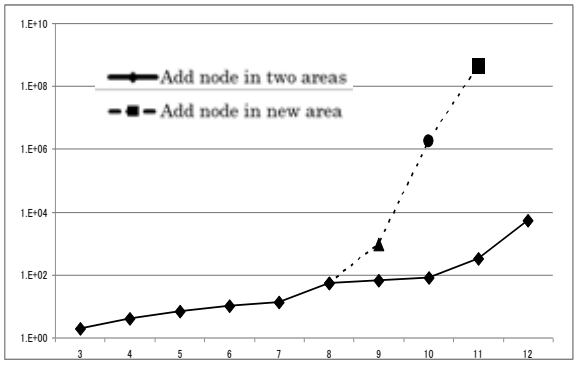


Fig.9: Total counts of route combinations.

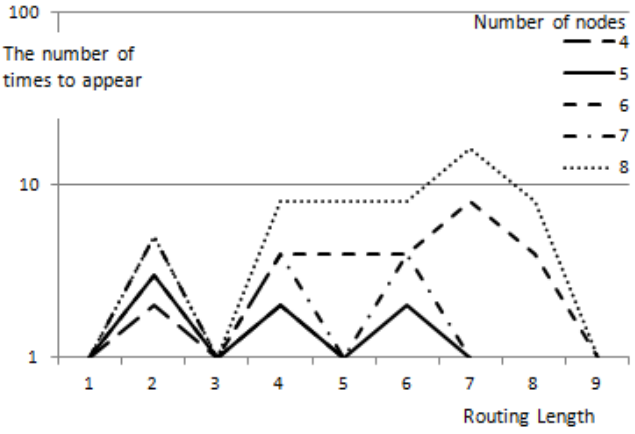


Fig.10: Histogram of simulation result: This histogram contains 4 nodes for 8 nodes. This histogram show less difference with number of nodes. So few nodes and less spread nodes can make mutually complementary network using wired and wireless communications.

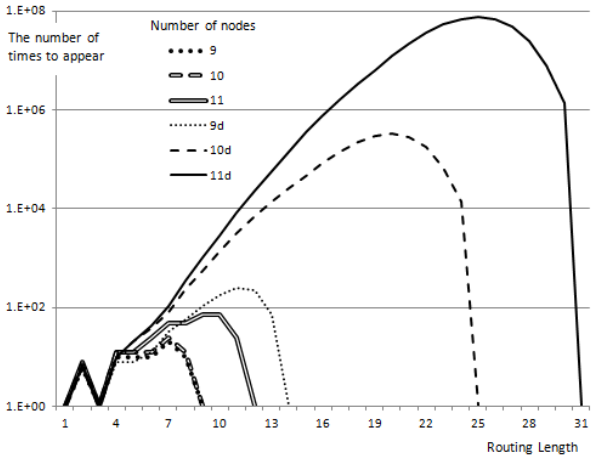


Fig.11: Histogram of simulation result: This histogram contains 9 nodes for 11 nodes. This result means a lot of nodes and wide spread nodes to need some roles to make mutually complementary network using wired and wireless communications.

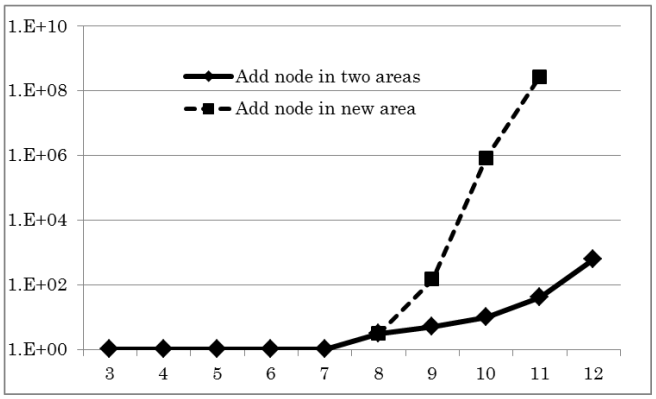


Fig.12: Time spent for simulation: Before optimization. The longest pattern spent over 80 hours.

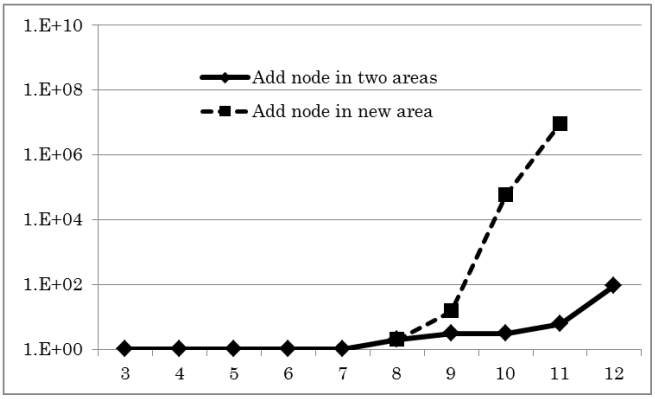


Fig.13: Time spent for simulation: After optimization. The longest time shortened for 3 hour.

Table.3 Results of routing simulation: eight nodes in a two-area building

SN	Route	adr	Tn	R	Try	TryAv
1	[1->2]	1->2	1	1	1	1
2	[1->3]	1->3	1	1	1	1
3	[1->4]	1->4	1	1	1	1
4	[1->5]	1->5	1	1	1	1
5	[1->(6)->2->6]	1->6	1		3	
6	[1->(6)->3->(6)->2->6]	1->6	2		5	
7	[1->(6)->3->(6)->4->6]	1->6	3		5	
8	[1->(6)->3->(6)->5->6]	1->6	4		5	
9	[1->(6)->3->(6)->7->6]	1->6	5		5	
10	[1->(6)->3->(6)->(8)->2->6]	1->6	6		6	
11	[1->(6)->3->(6)->(8)->4->6]	1->6	7		6	
12	[1->(6)->3->(6)->(8)->5->6]	1->6	8		6	
13	[1->(6)->3->(6)->(8)->7->6]	1->6	9		6	
14	[1->(6)->4->6]	1->6	10		3	
15	[1->(6)->5->6]	1->6	11		3	
16	[1->(6)->7->6]	1->6	12		3	
17	[1->(6)->(8)->2->6]	1->6	13		4	
18	[1->(6)->(8)->3->(6)->2->6]	1->6	14		6	
19	[1->(6)->(8)->3->(6)->4->6]	1->6	15		6	
20	[1->(6)->(8)->3->(6)->5->6]	1->6	16		6	
21	[1->(6)->(8)->3->(6)->7->6]	1->6	17		6	
22	[1->(6)->(8)->3->(6)->(8)->2->6]	1->6	18		7	
23	[1->(6)->(8)->3->(6)->(8)->4->6]	1->6	19		7	
24	[1->(6)->(8)->3->(6)->(8)->5->6]	1->6	20		7	
25	[1->(6)->(8)->3->(6)->(8)->7->6]	1->6	21		7	
26	[1->(6)->(8)->4->6]	1->6	22		4	
27	[1->(6)->(8)->5->6]	1->6	23		4	
28	[1->(6)->(8)->7->6]	1->6	24	24	4	5.167
29	[1->7]	1->7	1	1	1	1
30	[1->(8)->2->8]	1->8	1		3	
31	[1->(8)->3->(8)->2->8]	1->8	2		5	
32	[1->(8)->3->(8)->4->8]	1->8	3		5	
33	[1->(8)->3->(8)->5->8]	1->8	4		5	
34	[1->(8)->3->(8)->(6)->2->8]	1->8	5		6	
35	[1->(8)->3->(8)->(6)->4->8]	1->8	6		6	
36	[1->(8)->3->(8)->(6)->5->8]	1->8	7		6	
37	[1->(8)->3->(8)->(6)->7->8]	1->8	8		6	
38	[1->(8)->3->(8)->7->8]	1->8	9		5	
39	[1->(8)->4->8]	1->8	10		3	
40	[1->(8)->5->8]	1->8	11		3	
41	[1->(8)->(6)->2->8]	1->8	12		4	
42	[1->(8)->(6)->3->(8)->2->8]	1->8	13		6	
43	[1->(8)->(6)->3->(8)->4->8]	1->8	14		6	
44	[1->(8)->(6)->3->(8)->5->8]	1->8	15		6	
45	[1->(8)->(6)->3->(8)->(6)->2->8]	1->8	16		7	
46	[1->(8)->(6)->3->(8)->(6)->4->8]	1->8	17		7	
47	[1->(8)->(6)->3->(8)->(6)->5->8]	1->8	18		7	
48	[1->(8)->(6)->3->(8)->(6)->7->8]	1->8	19		7	
49	[1->(8)->(6)->3->(8)->7->8]	1->8	20		6	
50	[1->(8)->(6)->4->8]	1->8	21		4	
51	[1->(8)->(6)->5->8]	1->8	22		4	
52	[1->(8)->(6)->7->8]	1->8	23		4	
53	[1->(8)->7->8]	1->8	24	24	3	5.167

5. SUMMARY

For the study described herein, a mutually complementary network of wired and wireless communications was applied to a building such as a school; a simulator was developed to determine how effective a routing capability would be. Then its results were evaluated. Assuming that we can have only 10s available for communication knowing it based on the number of nodes considered, results show that up to nine nodes for a three-area building are just about the limit available for a practical use with a location-undefined communication method used. If 15s or less are assumed to be available, then it is up to 11 nodes for five-area building. Based on this evaluation to the case to apply the mutually complementary network for a school building, for example, we find it possible by allocating a domain unit of 11 nodes to the entire building.

Future studies to be made should first use a building with as few stories as possible and have a greater number of nodes under the associated communication duration time conditions given. Second, they should include that we have a network capability in which we have a domain unit of multiple nodes to allocate to the entire building. With a simulator used for the simulation to the setting of 11 nodes for a five-area building, the number of routing combinations that existed with our routing capability method exceeded 400 million, and the simulation duration conducted exceeded 70hr. in all. Even under those circumstances, the maximum number of routings performed was 29 times for up to 14.5s spent, which indicates that the routing method introduced herein can be available for practical use. It would be necessary in any case to reduce the simulation hours greatly.

Regarding mutually complementary networks incorporating wired and wireless technologies, it is important to have the right communication duration time and communication performance to make it available for a practical use, and to show a typical example of its use as well, thereby contributing to the security and the energy control for such a wide area that covers private homes and other buildings for which no proper communication method is available [7].

This time, by the function was separated into 2, the simulator sped up. For speed-up of the execution time of simulator, decrease the access frequency to HDD, and decrease the process clone of data creating. The latter is able to decrease memory used for simulator by using tree structure. By using these, aim for a practicable simulator, and develop of mutual complement network.

REFERENCES

[1] Kunihiro Yamada, etc. "Dual Communication System Using Wired and Wireless Correspondence in Small Space," KES2004, Springer LNAI3214, pp. 898-904, 2004

[2] M16C/6S http://www.renesas.com

[3] URL http://www.Zig-Bee.org

[4] Daiki Watabe, Kunihiro Yamada "Route simulation in mutually complementary network between wired and wireless," 2010Information Processing Society of Japan, pp. 3-313-3-314, 2010

[5] Kunihiro Yamada, Takashi Furumura, Yukihisa Naoe, Kinichi Kitazawa, Toru Shimizu, Kouji Yoshida, Masanori Kojima, Tadanori Mizuno "Home-Network of a Mutually Complementary Communication System by Wired and Wireless," KES2006, Springer Part III LNAI4253, pp. 189-196, 2006

[6] Incorporating the Best of Both Worlds for Improved Functionality Industrial Environments "Hybrid Wired/Wireless Network for Real-Time Communications," IEEE Industrial Electronics Magazine, March 2008, Vol. 2, No. 1, pp. 8-20

[7] Kunihiro Yamada " The home network by cooperation communication technology - mutually complementary network incorporating - "Intellectual property strategy headquarters.Industry-university cooperation technolog Tokai University.,
Collection of the research starting points.
Tokai University 2007 pp.29

Invited Speech: Dr. Hiroshi Shimodaira (Chair Kouji Yoshida)

HMM-based Speech-driven Talking Faces

Hiroshi Shimodaira
Gregor Hofer, Junichi Yamagishi

CSTR, Informatics, Univ. of Edinburgh

Motion Capture based Animatain

- **MoCap: widely used in film and game industry**

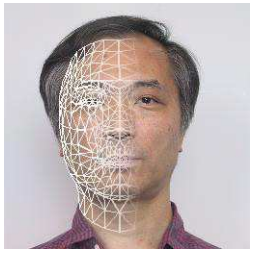
(Avatar, James Cameron, 2009)

Pros.
- **Good quality (accurate in time and space)**
- **Natural motion**
- **Could go beyond "uncanny valley"**

Cons.
- **Inflexible (difficult to modify/customise)**
- **Enormous time/work for recording**

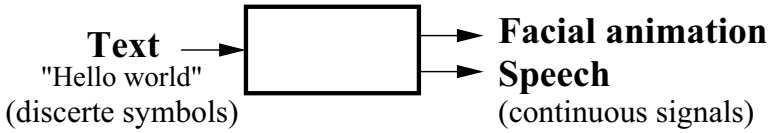
What is a speech-driven talking head?

It's a computer animated talking face, "talking head", whose motion is driven by speech input.

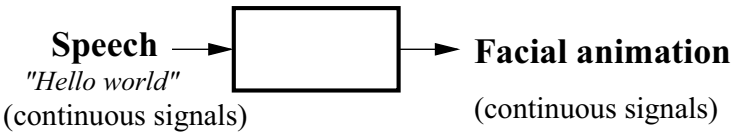


Two types of talking heads:

- **Text driven**



- **Speech driven**

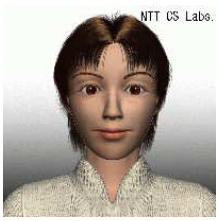


Applications

- **Film industries**
- **Computer games**
- **Agent-based system (spoken dialogue systems)**
- **Education (pronunciation training), psychotherapy**
- **Simulator for scientific research**

Examples of facial animation

August [J. Gustafson etal., KTH 1999]



Hyumu [Dosaka, NTT 2000]

Baldi [S. Sutton etal. OGI 1998]

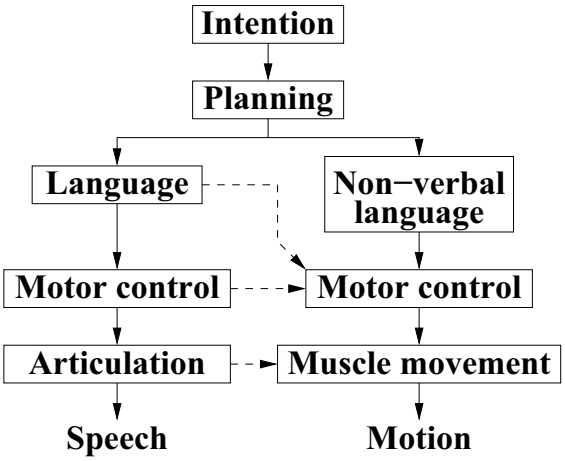
[Hasegawa etal., ETL 1999]

Mary101 [T. Ezzat etal., MIT 2002]

Greta [C. Pelachaud etal., Paris8 2000]

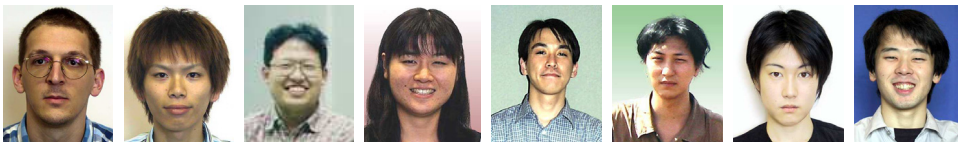
Controlling motion of talking heads

- Mouth/Lip motion (lip-sync)
- Eye motion (eye gaze, eye blinking)
- Eyebrow motion
- Head motion
- Facial expression



Why statistical approach is needed?

- Trainable on real data (of a specific real person). $\rightarrow$ Easy to create different agents as many as possible
- Adaptable to new speakers / styles
- Able to generate non-deterministic motions (a great deal of variety in motion)



Problem formulation

Define the problem as a probabilistic optimisation problem:

$$\boldsymbol{O}^{M*} = \arg\max_{\boldsymbol{O}^M} P(\boldsymbol{O}^M|\boldsymbol{O}^S)$$

$\boldsymbol{O}^S = \boldsymbol{o}^S_1, \boldsymbol{o}^S_2, \ldots, \boldsymbol{o}^S_{L^S}$ sequence of speech features

$\boldsymbol{O}^M = \boldsymbol{o}^M_1, \boldsymbol{o}^M_2, \ldots, \boldsymbol{o}^M_{L^M}$ sequence of motion features

- It's not a point-to-point mapping, but a stream-to-stream mapping of real-valued vectors, in which context should be taken into account. The mapping is usually not invertible.

Input \Output	Discrete	Continuous
Discrete	machine translation	text-to-speech synthesis
Continuous	speech recognition	(this problem)

Problem formulation *(cont. 2)*

- **Difficulty**
  - **The mapping seems to be complex, non-linear, context dependent.**
  - **Different POIs have different dependencies and different levels of synchrony on/with speech.**
  - **It's not clear what acoustic/language features and model unit should be used to predict motions of POI.**

POI	dependency on speech	literature
mouth & jaw	high	many
head	moderate?	several
eye (gaze, blink)	weak?	very few
eyebrow	weak?	very few

Simplified solution

Assume a model sequence: $\boldsymbol{u} = u_1, u_2, \ldots$

$$\boldsymbol{O}^{M*} = \arg\max_{\boldsymbol{O}^M} P(\boldsymbol{O}^M|\boldsymbol{O}^S) = \arg\max_{\boldsymbol{O}^M} \sum_{\boldsymbol{u}} P(\boldsymbol{O}^M, \boldsymbol{u}|\boldsymbol{O}^S)$$

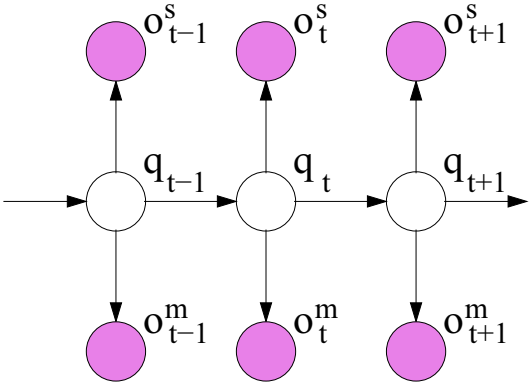
$$\approx \arg\max_{\boldsymbol{O}^M} \sum_{\boldsymbol{u}} P(\boldsymbol{O}^M|\boldsymbol{u}) P(\boldsymbol{u}|\boldsymbol{O}^S)$$

$$\approx \arg\max_{\boldsymbol{O}^M} P(\boldsymbol{O}^M|\boldsymbol{u}^*)$$

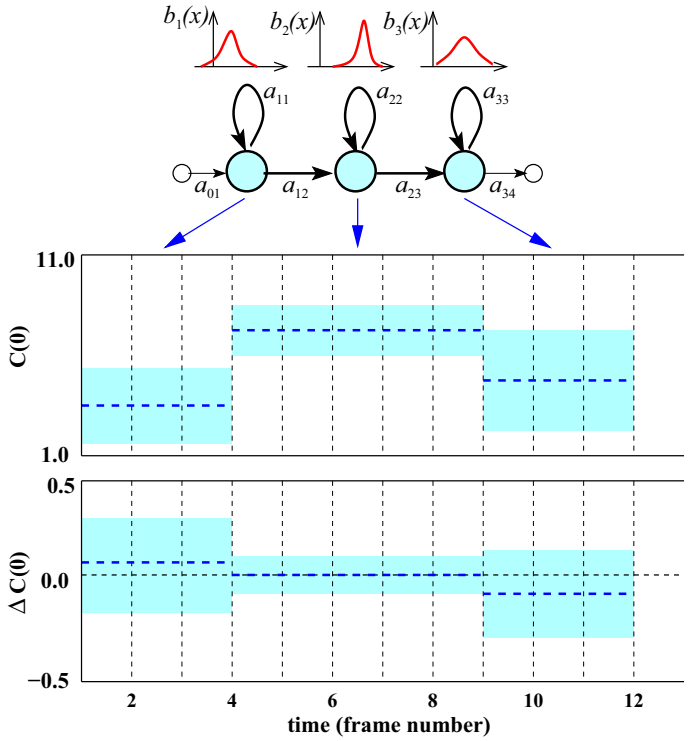
$$\boldsymbol{u}^* = \arg\max_{\boldsymbol{u}} P(\boldsymbol{u}|\boldsymbol{O}^S)$$

This is referred as "HMM re-mapping".

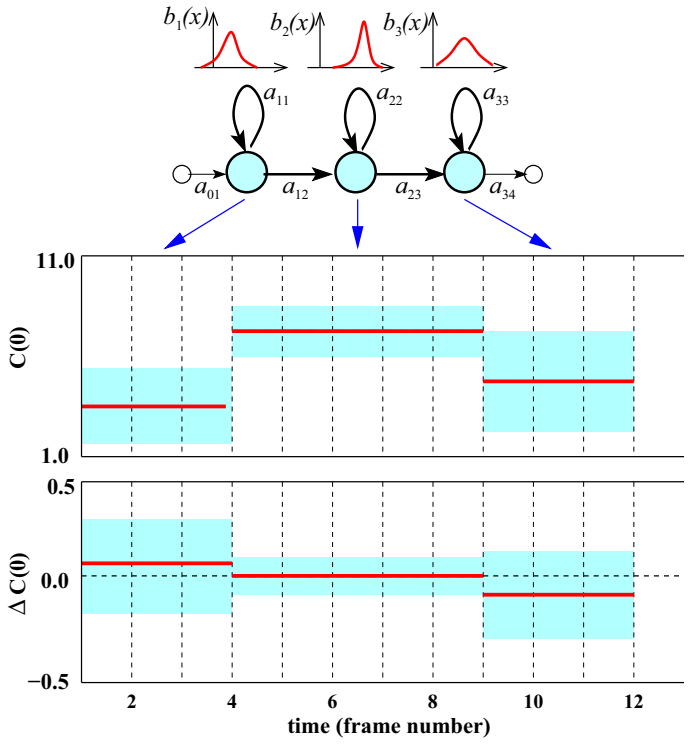
Matthew Brand, "Voice Puppetry", 1999



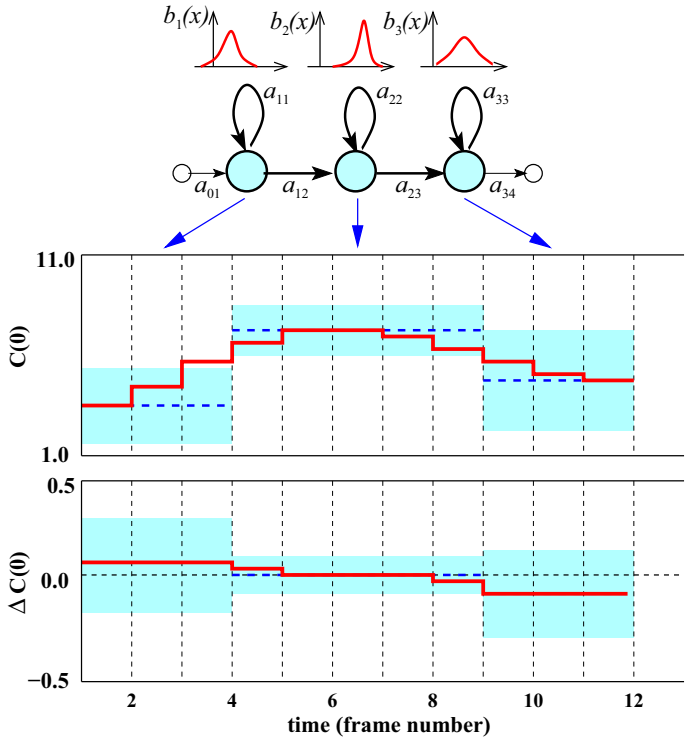
Generating signals from a HMM



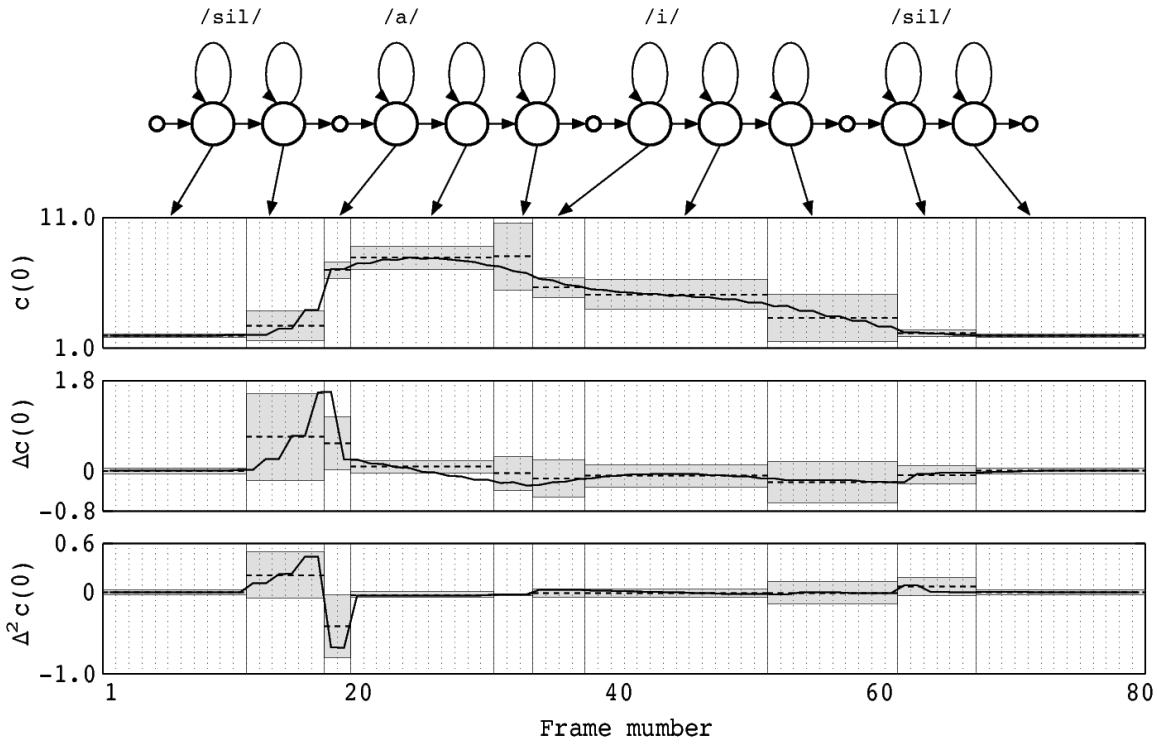
Generating signals from a HMM



Signal synthesis from trajectory-HMM



Signal synthesis from Trajectory-HMM



Trajectory HMM (K. Tokuda etal, 1995, 2007)

$$\boldsymbol{O} = [\, \boldsymbol{o}_1, \ \ldots, \ \boldsymbol{o}_T \,] = \begin{bmatrix} \boldsymbol{C} \\ \Delta \boldsymbol{C} \\ \Delta^2 \boldsymbol{C} \end{bmatrix} = \begin{bmatrix} \boldsymbol{c}_1, & \ldots, & \boldsymbol{c}_T \\ \Delta \boldsymbol{c}_1, & \ldots, & \Delta \boldsymbol{c}_T \\ \Delta^2 \boldsymbol{c}_1, & \ldots, & \Delta^2 \boldsymbol{c}_T \end{bmatrix}$$

$$\boldsymbol{c}_t = [c_t(1), \ldots, c_t(M)]^T$$

NB: O is derived from C, i.e. $\tilde{O} = W\tilde{C}$

$$\begin{aligned} p(\boldsymbol{O}|\Lambda) &= \sum_{\boldsymbol{q}} p(\boldsymbol{O}|\boldsymbol{q}, \Lambda)\, p(\boldsymbol{q}|\Lambda) \\ &= \sum_{\boldsymbol{q}} \prod_{t=1}^{T} p(\boldsymbol{o}_t|q_t, \Lambda)\, p(q_t|q_{t-1}, \Lambda) \end{aligned}$$

standard HMM : $\boldsymbol{O}^* = \arg\max_{\boldsymbol{O}} P(\boldsymbol{O}|\Lambda)$

trajectory HMM : $\boldsymbol{O}^* = \arg\max_{\boldsymbol{O}} P(\boldsymbol{O}|\Lambda)$ w.r.t. $\tilde{\boldsymbol{O}} = W\tilde{\boldsymbol{C}}$

(K. Tokuda etal, 1995, 2007)

Head motion synthesis

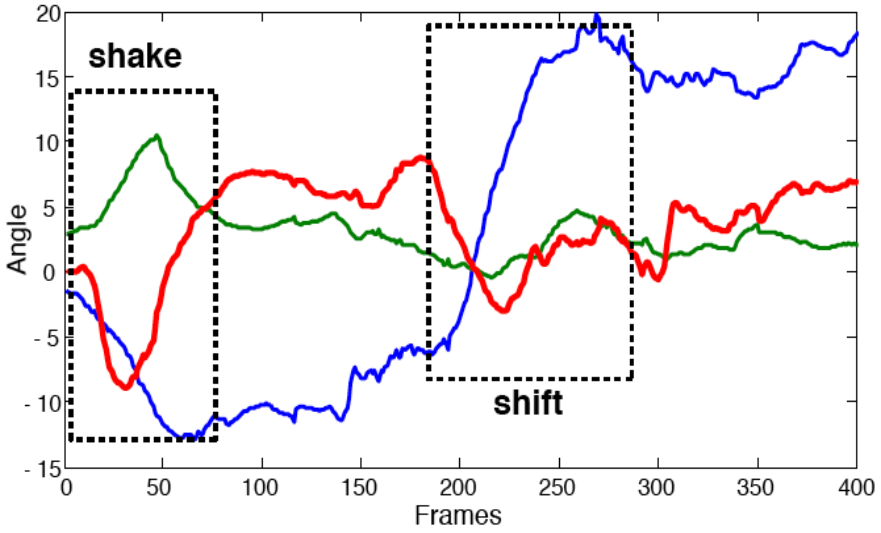
- **Possible units and features**

Domain	Feature	unit	auxiliary labels
speech	text acoustic	phoneme/syllable word (prosodic) phrase	accent, loudness POS, position pause
head motion	direction (angles)	manual clustering	

- **Selected unit: 4 types of head motions**

postural shift : the head shifts axis of movement
shake & nod : lateral movement around one axis
pause : no movement / rest position
default : non-distinctive movement or slow movement

Example of motion unit

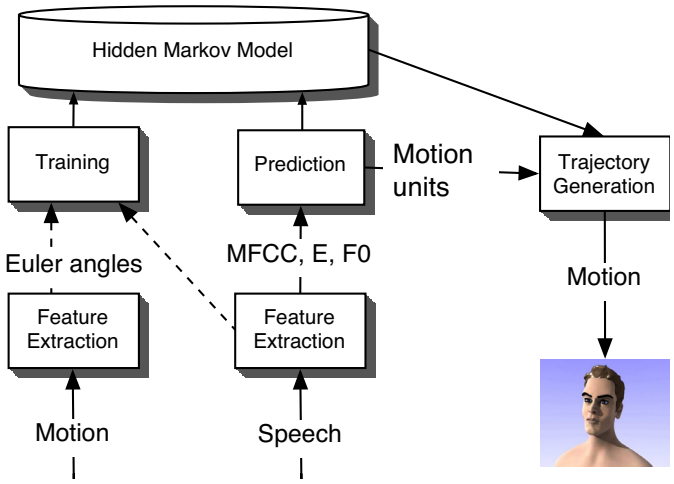


	shift	shake	pause	default
# samples	209	107	47	209
avg. length [s]	0.4	0.8	0.8	25

Data collection

- Motion Capture with reflective markers (Qualisys)
- 2 English male native RP speakers
  - 25 minutes of free speech
  - 500 utterances read speech
- 5 speakers with various accents
  - only 10 minutes of free speech

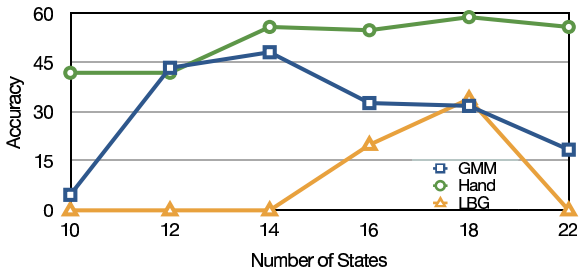
Training and synthesis



- **Training**
  - **Train HMMs with the two-stream date (speech + motion)**
- **Synthesis**
  - **Recognise the input speech with the HMMs**
  - **Synthesise trajectories of head motion (trajectory HMM)**

Comparison of unit types

Automatic Labels



- Comparison of automatically determined labels with manually determined labels
  - clustering into 4 classes: LBG and Gaussian Mixture
  - consecutive frames are grouped as one sequence

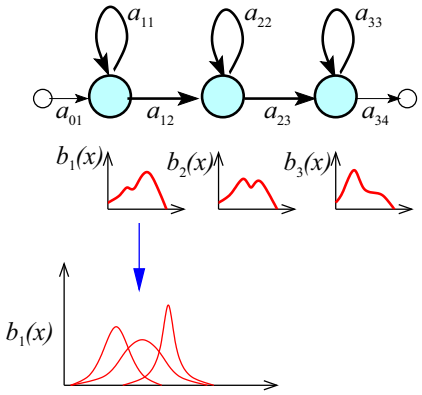
Extension of trajectory-HMM

Problems with HMM-based motion synthesis

- **Output has a narrower dynamic range**
- **Output is deterministic**

Possible solutions

- **put a global variance term into the objective function.**
- **select a mixture component of GMM of each state randomly.**



Subjective evaluation

Cartoon BW	Cartoon Color	Plain	Textured
			
Black-and-white cartoon shader render	Color cartoon shader render	Single colour (beige) ray tracing render	Textured ray tracing render

Related works

- Lip-synchronisation
  - HMM: S. Nakamura (ATR, 2001), C. Busso and U. Neumann (USC, 2007), [HMM-inversion] K.Choi and J-N Hwang (U.Washington,1999)
  - AHMM-CHMM-VHMM: P.S. Aleksic and A.K. Katsaggelos (Northwestern U., 2004)
  - DBN: L. Xie and A-Q Liu (U. Hong Kong, 2007)
  - Trajectory-HMM: K.Tokuda (Nagoya Inst., 2000)
  - Switching linear dynamical systems: G. Englebienne and T. Cootes (Manchester U., 2008)

Related works *(cont. 2)*

- Head motion synthesis
  - HMM clustering: E. Sargin (Koc U. 2007)
- Eye motion synthesis
  - GMM: S.P. Lee, J. Badler, N. Badler (2002)

Summary

- Speech-driven head-motion synthesis with trajectory-HMM
- Human-readable model unit rather than automatically-determined unit.
- Further extensions
  - efficient use of prosody
  - incorporate language information

Session 4: Network2
(Chair Yoh Shiraishi)

Access-Based Contents Grouping on P2P Network

Takuya Sasaki[*], Jun Sawamoto[*], Takashi Katoh[*], Yuji Wada[**], Norihisa Segawa[*] and Eiji Sugino[*]

[*] Faculty of Software and information Science, Iwate Prefectural University, Japan
[**] Department of Information Environment, Tokyo Denki University, Japan
g231i017@s.iwate-pu.ac.jp, sawamoto@iwate-pu.ac.jp, t-kato@soft.iwate-pu.ac.jp, yujiwada@sie.dendai.ac.jp, sega@iwate-pu.ac.jp and, sugino@iwate-pu.ac.jp

Abstract –In recent years, the diversification and ubiquitousness of information is rapidly advancing due to the development of the information and communication technology. However, as for the management and usage of those information sources, it is the usual case that individual and independent database management system is used for managing and maintaining each information source. As for the usage of the information, the attention of sharing of information using P2P network rises, and it is becoming to be used in many fields. The distributed hash table (DHT) is one of the typical overlay networks on P2P. However, there is a problem of lacking of flexibility in the retrieval on DHT. In this paper, to improve the convenience of the content access, we propose an efficient content access method based on the grouping technique of the contents using the content access history and frequency and the evaluation of grouping efficiency (precision, recall) is performed by the simple simulation.

Keywords: P2P network, distributed hash table, overlay network, content retrieval, grouping technique.

1 INTRODUCTION

In recent years, for the user who acts while moving around with a mobile device, the free mobile computing environment with an ad hoc communication like wireless LAN and the short distance wireless access are coming to be achieved and removing the restraint on user's movement [1].

As for the usage of the information, the attention of sharing of information using P2P network rises, and it is becoming to be used in many fields [2], [3], [4]. P2P networks directly connect computers in an equal relationship and the load is distributed by exchanging information and services among computers, and high fault tolerance is also one of the advantages. P2P networks can be classified into two types; unstructured and structured overlay network, by the transfer method of the search messages. Flooding is a fundamental search method in unstructured P2P systems [5], [6]. However, the searching cost is high by generating a huge amount of network traffic. Structured P2P network employ a globally consistent protocol to ensure that any node can efficiently route a search to some peer that has the desired content. The most common type of structured P2P network is the DHT [7], [8], in which a variant of consistent hashing is used to assign ownership of each content to a particular peer. Therefore, DHT can search content with the number of very few messages in comparison with the flooding method; however, the perfect matching with the search keyword is demanded in the hush table. The DHT has the drawbacks that the partial matching or the group matching retrieval cannot be achieved. It is difficult to do a flexible retrieval. This becomes a problem that should be solved when the search engine is constructed.

In this paper, to improve the convenience and flexibility of the content retrieval, we propose an efficient content access method based on the grouping technique of the contents using the content access history and frequency in ubiquitous P2P network environment. We introduce grouping structure into DHT in P2P and enhance conventional keyword search capability. In addition, we consider mechanism of automatic generation and reformation of group structure based on the users' access records. In DHT, it is a burden for the network manager to renew the hush table when a new group appears or an old group is deleted. Then, the automatic grouping is examined in this paper. When a user accesses the system and retrieves data, we can consider a set of continuous data retrievals as a candidate for grouping. When not only one user but also multiple users are doing a similar set of retrievals, they could be registered in the DHT as a new group. At the same time, the group is updated in proportion to the retrieval frequency. Less frequently retrieved groups are deleted from the table. Thus, the user can find favorable data more easily and efficiently. Finally, we evaluate the proposed method by simulation. In this study, it is targeted to simulate and to evaluate whether the automatic grouping of the DHT is carried out as we assumed in P2P network environment.

2 CONTENT GROUPING IN P2P NETWORK

P2P overlay networks are distributed systems without any hierarchical organization or centralized control. DHT-based systems can guarantee that any data object can be located in a small O(logN) overlay hops on average, where N is the number of peer nodes in the system.

P2P overlay networks are distributed systems without any hierarchical organization or centralized control. Such Structured P2P systems use the DHT as a substrate, in which data object (or value) location information is placed deterministically, at the peers with identifiers corresponding to the data object's unique key. Keys are mapped by the overlay network protocol to a unique live peer in the overlay network. The P2P overlay networks support the scalable storage and retrieval of {key, value} pairs on the overlay network. Given a key, a store operation (put(key, value))

lookup retrieval operation (value=get(key)) can be invoked to store and retrieve the data object corresponding to the key. DHT-based systems can guarantee that any data object can be located in a small O(logN) overlay hops on average, where N is the number of peer nodes in the system. However, DHT is lacking of flexibility in retrieval of plural contents at the same time, or in the partial-match retrieval by the keyword in P2P overlay networks.

On the other hand, in unstructured P2P networks, flooding or random walk is the predominant search technique [5]. The unstructured network topology is highly robust to failures or node transience. It is also relatively straightforward to implement multiple-keyword retrieval or partial match. On the down side, flooding causes high volumes of network traffic overheads, and the search results are nondeterministic. The performance of search in unstructured P2P overlay networks can be improved by using smarter search or data replication algorithms [6].

2.1 Contents grouping method

In this paper, we propose an automatic content grouping method in the P2P network by considering the continuity of the retrieval actions by the user. Generally, the cost to renew the hush table in DHT is high. Churning of the node occurs frequently because ad hoc participation and withdrawal of node is assumed. In addition, the frequency in which the DHT is renewed increases if it comes to group contents according to user's access results frequently. Then, we propose an automatic grouping function of the DHT. By this function, it becomes possible to offer contents grouping and renewal of DHT in real time.

In this paper, the Kademlia algorithm [8] is used for the DHT. As for the Kademlia algorithm, it is evaluated as a strong algorithm for the ad hoc node participation and secession. The accuracy of the contents retrieval decreases if DHT doesn't update according to the change of nodes. Though updating DHT in real time is ideal, the updating process requires user's work. Then, we propose a method of automatic generation of new contents group and deletion of old contents group according to user's retrieval actions.

It is possible to view those accessed contents as one group when a certain user uses the system, and user retrieves not only a single content but multiple contents continuously. The group of contents that a certain user used has the possibility to be used as a useful content group for other users. When not only the user but also another user is doing a similar retrieval, they could be registered in the DHT as a new group. However, when a flow of time and a new kind of item appears, the item that comes off as a retrieval target is not permanently left as a member of a group but the group is updated in accordance with the access frequency etc. enabling the access efficient and suitable for tendencies such as the taste of users or the fashion on the net in real time.

2.2 Related work

In this section, we discuss about related studies about the grouping of contents in the internet environment. In the case of Web pages, Tajima et al. [9] propose to group them from the similarity degree of the characteristic vectors of Web pages. Avan et al. [10] examine the grouping that uses the meta-data of the Web page, and Mizuuchi et al. [11] looks for an important page from link relations on the Web. In the study on the grouping of the Web community, Kaneko et al. [12] extract human relations of the specific community automatically only from information on the Web using Web mining technology, and Murata [13] discovers Web communities based on the co-occurrence of references. Yenta [14] intermediates between people with the same taste using a mobile agent, and performs grouping by the relevance of the taste.

There are a lot of studies of grouping as shown above even if the objects of the grouping differ. The object of the grouping in this paper is the contents that nodes of P2P network maintain. Grouping of nodes (contents) of P2P network is based on dynamic access history and access frequency by user nodes because nodes participate and leave the network frequently. In addition, the discovery of new group of contents and the timeliness that can reflect tendencies such as the taste of the user or the fashion on the net in real time are strong characteristics to update the group whenever an access is performed.

2.3 Relevance index for contents grouping

A P2P node works as a content holder and at the same time a user node who executes accesses to the contents on other nodes. When a user accesses plural contents in succession within a certain time period, e.g., a day, we assume that those contents become to have high relevance indices with each other. Relevance index shows how much value the continuous access of two contents has for users. Relevance index is shown by equation (1). Here, $K_{s,i}$ shows the relevance index between contents *s* and *i*. $v_{s,i}^{n}$ has value 1 if contents *s* and *i* are continuously accessed by a user at P2P node *n*, otherwise has value 0. In other words, Relevance index increases by value 1 when continuous access takes place by a user at some P2P node.

On the other hand, because time passes, in a contents group which has been formed, and was not used, relevance index decreases, and the grouping is broken off. The rate of reduction of the relevance index reflects the speed by that less frequently accessed group contents lose their relevance as a group.

$$K_{s,i} = \sum_{n \in nodes\ in\ P2P} v_{s,i}^{n} \qquad (1)$$

2.4 A small example of group creation and deletion (example of PC parts)

An example flow of the automatic creation of a new group that occurs when users actually access and contents are downloaded, and deletion of a group when it becomes not used, is described as follows and shown in Figure 1.

1. Contents registered in the DHT do not belong to a group in the initial state. Contents are shown in (a). "P6T" is a name of a mother board, "CD-552GA" is a CD drive and "Core i7 920" is a CPU respectively.

2. These three contents are retrieved and downloaded frequently by many users.
3. These three contents are registered in the DHT as a group accessed often at the same time. Grouped contents are shown in (b).
4. DVD drive appears. The DVD drive is a content named "DVR-216DBK" here. The state is shown in (c).
5. The user comes to do the retrieval of the combination of DVD drive, CPU, and mother boards by the appearance of the DVD drive in place of the group that contains the CD drive.
6. The group is newly generated because the combination of DVD drive, CPU, and the mother board frequently came to be retrieved by the user, and it is registered to the DHT. (d) shows the state in which a group is newly generated.
7. The group including the DVD drive comes to be retrieved and downloaded by more users than the group of the CD drive. The CD drive after the DVD drive appears is retrieved and downloaded less frequently.
8. The group that contains the CD drive enters the state that is with less profitable information for the user. Then, the group that contains the CD drive is deleted as shown in (e). However, only the group of the CD drive is deleted, and CD drive itself is possible to be retrieved and downloaded as a single content.

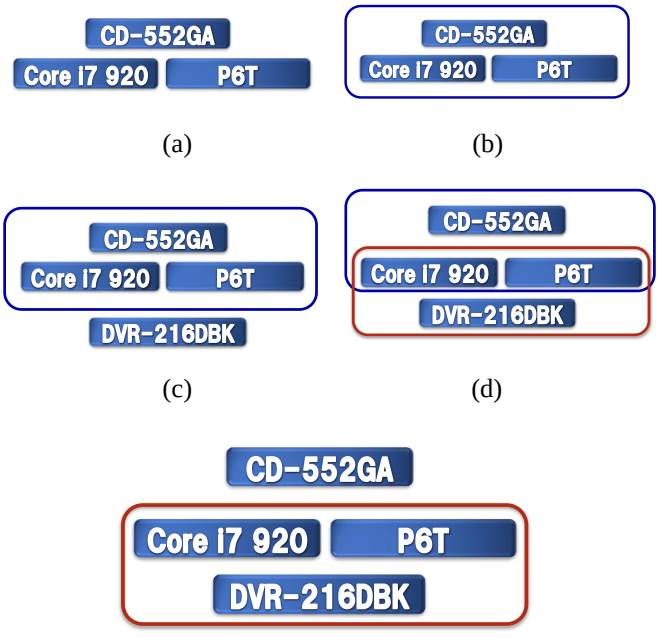


Figure 1: A small example of group creation and deletion: (a) Initial state without a group, (b) A group is generated, (c) Appearance of DVD drive, (d) New contents group is generated, (e) The group that contains the CD drive is deleted.

3 SYSTEM CONFIGURATION

P2P network is constructed with the DHT that uses Kademlia algorithm [8]. A P2P node usually participates to the network as a content holder (provider) and/or a user (consumer) of contents as shown in Figure 2. In this system, we assume that one node holds a single content for simplicity. However, it is an easy extension to make a P2P node hold multiple contents in this configuration.

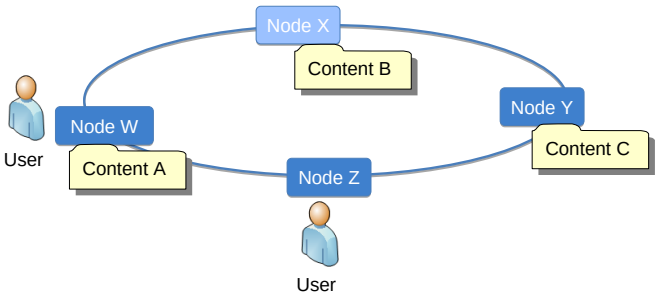


Figure 2: P2P network configuration where a node usually participates to the network as a content holder (provider) and/or a user (consumer) of contents.

3.1 P2P node

In general, the node of P2P network holds hash table and the contents information. In this paper, a list of access history and a group list are hold at each node. Figure 3 shows the state that keeps "a list of access history", "a group list" in a P2P node.

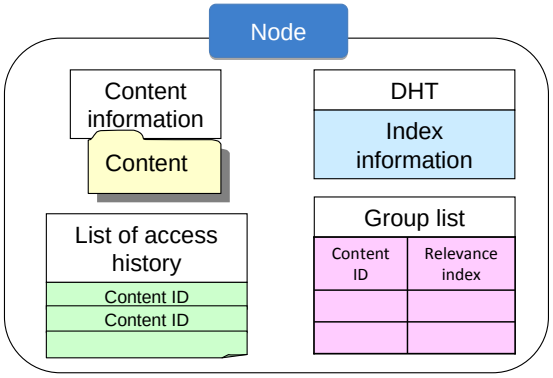


Figure 3: P2P node keeps "a list of access history", "a group list" in addition to hash table and the contents information.

Following four data structures are located in a P2P node.

1) *Contents information:* Content files which this node holds.
2) *Distributed hash table:* Index information of the node, i.e., hash value and node ID, are hold and updated.
3) *A list of access history:* We maintain a list of access history for a set of continuous access actions at a node. When there is a node which holds the object content that a user is accessing, the content ID is added to the list of access history, and it is erased when the acquisition of contents are completed.
4) *A group list:* We maintain a group list which consists of content ID of high relevance index with this

hosting node, and it is sorted in descending order of the relevance index value. As time passes, the index values in the group list are maintained and decreases. When this hosting node is accessed from a user node, contents with high relevance indices are recommended back to the user node.

3.2 Contents grouping by content access history and frequency

The communication process between nodes for the grouping of the contents (or here, nodes which contain contents) is detailed (Figure 4). Below, we describe the communication process step by step.

① A user at node A requests to access content at node B after some retrieval action.

② Node A sends its own list of access history to node B. At node B, after receiving the list of access history, contents in the received list are merged to the group list of node B. If contents with the same IDs exist already in the B's group list, only relevance indices are increased by one.

③ Node B returns the requested own content to node A.

④ Node B sends back information of contents, which are members of B's group list and have relevance indices higher than a certain threshold set by the system parameter, to node A.

⑤ At node A, B's content IDs are added to the access history. By using information of contents from node B, A might access contents in the B's group list. This function is called the recommendation of contents in terms of the value of relevance index by node B to node A.

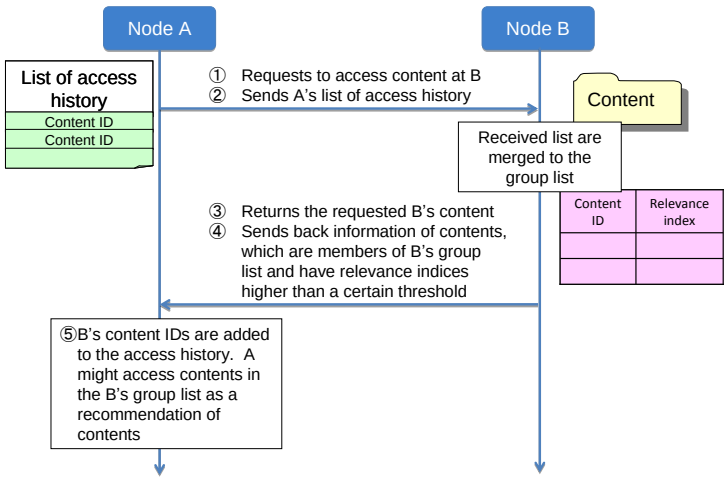


Figure 4: The communication process between nodes for the grouping of the contents by content access history and frequency.

3.3 System parameters for contents grouping

Table 1 shows system parameters that are main factors for determining characteristics of contents grouping. Number of participating nodes and number of nodes which hold contents determine the size of P2P network. In this situation, nodes are classified into two types; node with or without content. A node without content mainly accesses contents as a user. Access tendency index and access count are considered to affect how quickly contents groups are formed in the process. The rate of reduction of the relevance index reflects the fact that less frequently accessed group contents lose their relevance in a group.

Table 1: System parameters for contents grouping

Name of parameter	Explanation
Number of nodes	Maximum number of participating nodes in P2P network.
Number of content nodes	Number of nodes which hold contents.
Max size of the list of the access history	Maximum number of content items the list of access history holds. The higher the value, the more it is possible to maintain the user's access history. It is set to hold the max number of accesses in one access cycle of an average user.
Max size of the group list	Maximum number of content items the group list hold. This parameter determines the size of the group that can be formed in the system. Number of highly relevant contents increases by increasing this size.
Reduction rate of the relevance index	The rate the relevance index is reduced in one day period. If the value is x% then current relevance index S becomes S*x/100.
Threshold of relevance index	The threshold to tell whether contents are considered in the same group or not by the relevance among contents.
Access tendency index	Access tendency index indicates how frequently the user accesses contents which are candidates for grouping in a short period, e.g. a day. The tendency is higher the group could be formed more quickly.
Average access frequency	The average number of accesses a user executes a day on average.
Recommendation function (on/off)	This function is a recommendation function of contents in terms of the value of relevance index. If recommendation function is on, it could be considered to accelerate formation of contents grouping.

4 PERFORMANCE EVALUATION

We use the Distributed Environment Emulator of "Overlay Weaver" [15], which is an overlay construction toolkit, and perform the experiment. We prepare the scenario file for the emulator which appointed the movement of the node such as retrieval /acquisition of contents.

The typical flow of the scenario file is shown in Figure 5. For the content node we pre-set temporary groups beforehand in this scenario. The access of contents by user node is carried out from the pre-determined candidate

groups based on the access tendency index. Steps D) to G) correspond to one day activity at each node. At the end of the day, the relevance index is reduced by the reduction rate given by the parameter. In this experiment, we iterate this activity for thirty days. We run experiments for the combination of parameter values and calculate performance measures for output data and then evaluate the performance of the grouping process.

A) Generate participating P2P nodes
B) Construct overlay P2P network with generated nodes
C) Register nodes that hold contents (content nodes) to DHT
D) Each node execute retrieval by keyword
E) Each node access and acquire retrieved content
F) Relevance indices are reduced at each content node
G) Go to step D)

Figure 5: The typical flow of the scenario file

Table 2 shows parameter values used for the experiment. We run experiments according to the scenario for the combination of parameter values and calculate performance measures based on the output data and then evaluate the performance of the grouping process.

Table 2: Parameters set for the experiment

Name of parameter	Parameter values
Number of nodes	10000, 2000 nodes
Number of content nodes	1000 nodes
Max size of the list of access history	5
Max size of the group list	10
Reduction rate of the relevance index	5%, 10%, 30%, 50%
Threshold of relevance index	1, 2, 3
Access tendency index	20%, 50%, 80%
Average access frequency	3, 6, 9
Recommendation function (on/off)	on, off

4.1 Performance measures

The measures we use to express the performance of contents grouping are:

1) *Number of content nodes that are in the formed groups:* Number of content nodes that have high relevance indices with intended candidate nodes and can be considered to be forming a group. Number of content nodes is 1000 and 1000 is the maximum value.
2) *Average number of contents that are in the group lists:* Average number of contents that are in the group lists of nodes which are members of the group. Maximum size of group list is 10 and 10 is the maximum number.
3) *Precision:* It is the fraction of the contents in the correctly (as intended) formed group relevant to contents in formed group. Precision is shown in equation (2), where, P: precision, k: number of contents in the correct groups, n: number of intended groups of contents, m: number of formed group, l: maximum size of group list = 10.

$$P = \frac{\sum_1^n (\frac{k}{l})}{m} \quad (2)$$

4) *Recall:* It is the fraction of the contents in the correctly formed group relevant to all contents. Recall is shown in equation (3), where, R: recall, k: number of contents in the correct groups, n: number of intended groups of contents, l: maximum size of group list = 10.

$$R = \frac{\sum_1^n (\frac{k}{l})}{n} \quad (3)$$

4.2 Experimental results and consideration

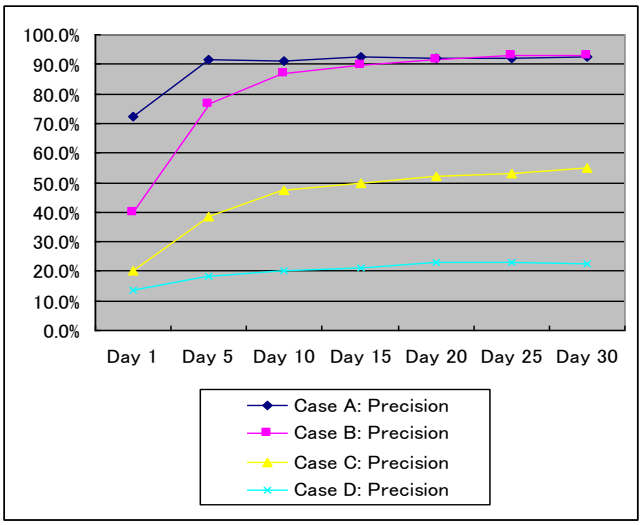
In this section, we explain some of the experimental results and discuss the feature of the proposed contents grouping method.

4.2.1 Effect of number of nodes

This experiment is comparing how grouping changes by the number of participating nodes. We evaluate it using parameter values of Case A through D as show in Table 3. Figure 6 is the graphs which express precision and recall for thirty days of the experiments. We examine the group formation process by increasing the participating nodes to 2,000 and 10,000. Grouping is performed more rapidly by increasing nodes, and both precision and recall show very high values. It can say number of participating nodes increases in P2P network, the proposed method for contents grouping shows a very high effect.

Table 3: Parameters set for the experiment

Cases	No. of nodes	Max size of the access history	Reduc. rate of the relev. index	Threshold of relev. index	Access tendency index	Ave. access freq.	Recomm. function
A	10000	10	5%	2	80%	3	on
B	10000	10	5%	2	50%	3	on
C	2000	10	5%	2	80%	3	on
D	2000	10	5%	2	50%	3	on



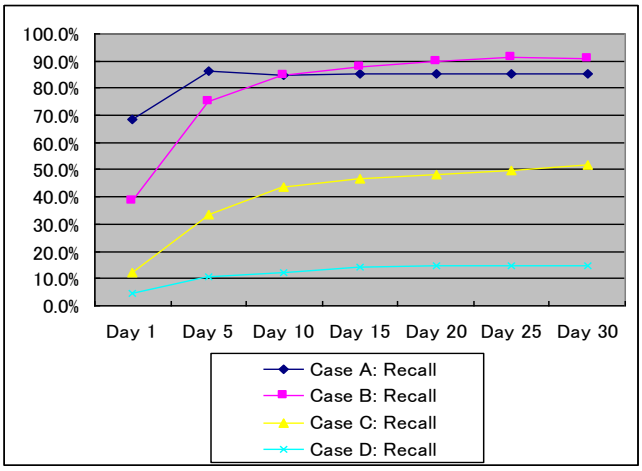


Figure 6: This experiment is comparing how grouping changes (precision and recall) by the number of participating nodes

4.2.2 Effect of the recommendation function

This experiment investigates the effect of the recommendation function. We evaluate it using parameter values of Case E through J as show in Table 4.

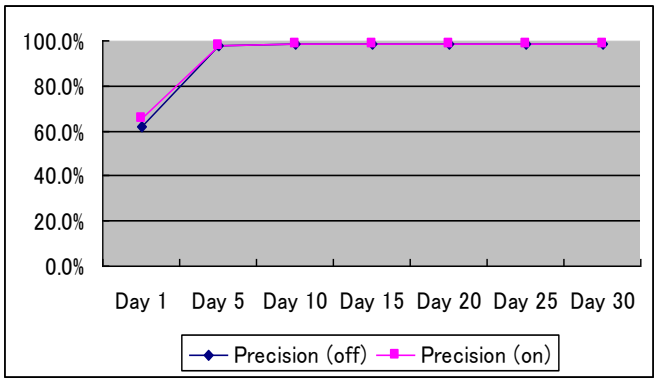
Table 4: Parameters set for the experiment

Case	No. of nodes	Reduc. rate of the relev. index	Threshold of relev. index	Access tendency index	Ave. access freq.	Recomm. function
E	10000	5%	3	80%	3	off
F	10000	5%	3	80%	3	on
G	10000	5%	3	50%	3	off
H	10000	5%	3	50%	3	on
I	10000	5%	3	20%	3	off
J	10000	5%	3	20%	3	on

In this experiment, it is shown that the recommendation function works for certain range of access tendency index. We can confirm that the recommendation function accelerates the formation of groups in the daily retrieval and access activities of users on P2P network environment.

(1) In case the access tendency index is 80% (Cases E, F)

Figure 7 shows the outcome of experiments as precision and recall value where the access tendency index is 80% (Cases E and F in Table 4). In Figure 7, keywords on/off indicate whether the recommendation function is used or not. The result shows that under this circumstance, both precision and recall are very high (almost 100%) after day 5 irrelevant of the recommendation function. Access tendency index indicates how frequently the user accesses contents which are candidates for grouping in a day. Therefore, in case of access tendency index 80%, the grouping of contents is achieved very quickly without introducing recommendation function.



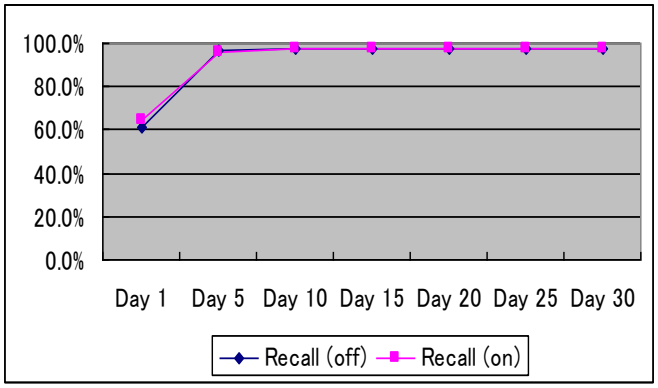


Figure 7: Precision and recall in case the recommendation function is on/off where the access tendency index is 80%

(2) In case the access tendency index is 50% (Cases G, H)

Figure 8 shows the outcome of experiments as precision and recall value where the access tendency index is 50% (Cases G and H in Table 4). In this experiment, both precision and recall are decreased drastically compared to cases E and F (access tendency index is set to 80%) when the recommendation function is off. However, with the recommendation function on, both precision and recall reach 90% after day 15. This result shows that even if users access irrelevant contents 50% in one access cycle, it is possible to choose necessary contents and form a group with the help of

recommendation function. In this case, we could confirm the usefulness of the recommendation function.

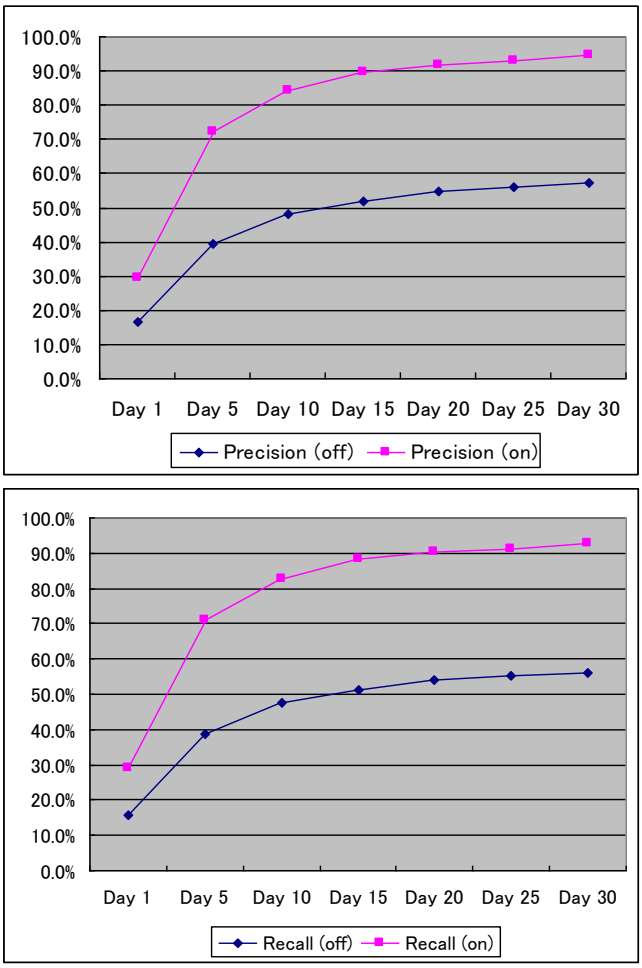
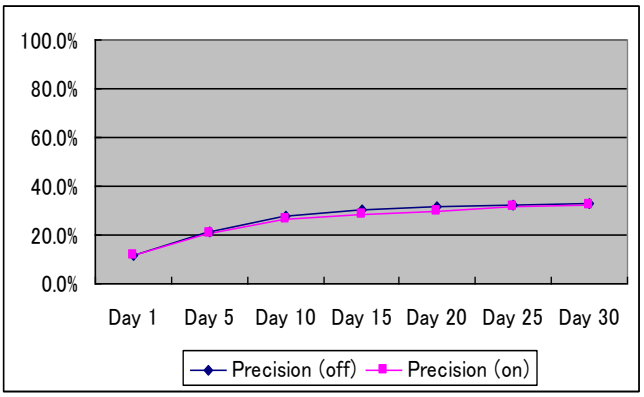


Figure 8: Precision and recall in case the recommendation function is on/off where the access tendency index is 50%

(3) In case the access tendency index is 20% (Cases I, J)

Figure 9 shows the outcome of experiments as precision and recall value where the access tendency index is 20% (Cases I and J in Table 4). In this case, users access almost irrelevant contents and even with the recommendation function on, meaningful grouping is not performed and both precision and recall are very low.



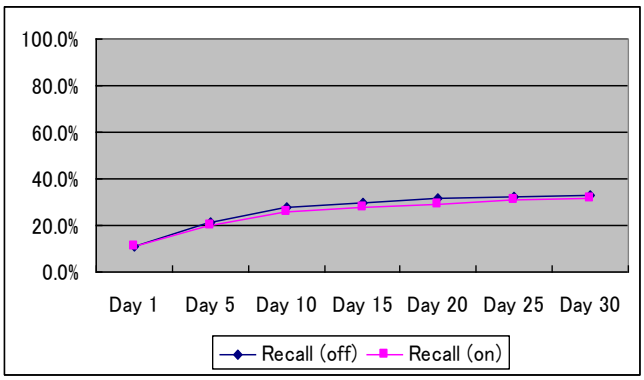


Figure 9: Precision and recall in case of the recommendation function is on/off where the access tendency index is 20%

4.2.3 Consideration

In this experiment, we fixed the number of P2P nodes and verify the basic feature of the proposed access-based contents grouping function. And we confirmed the effect of the recommendation function under certain noisy access behavior of the user. Because this function forms a contents group dynamically depending on the content access request issued from users, it can be said that the flexible feature has been implemented that corresponds to dynamic changes of nodes, i.e., ad hoc join and secession, in P2P network. By providing a content group that an interested content belongs when a certain search request is made as we have shown in this paper, we assume that the efficiency of the contents retrieval action is improved to some extent.

5 CONCLUSION

In this paper, we proposed an access-based contents grouping method aiming at an information retrieval system targeting contents in the P2P network, and performed evaluation by simulation. We intend to implement and evaluate the full function of grouping of contents in DHT. Based on the evaluation, improving the function of grouping of contents, it is scheduled to go to the system design of the efficient ubiquitous data retrieval system. In addition, the verification of the usability is scheduled to be covered in consideration of a concrete application system, and experimenting with the realistic applications.

ACKNOWLEDGMENT

This research was supported by a Grant-in-Aid for Scientific Research (C) (Subject No. 20500095) and Grant-in-Aid for Young Scientists (B) (Subject No. 21700084).

REFERENCES

[1] Chakraborty, D., et al.: Toward Distributed Service Discovery in Pervasive Computing Environments. IEEE Transactions on Mobile Computing, Vol. 5, No. 2, pp. 97–112 (2006).
[2] “Napster”, http://www.napster.jp/.

[3] “Gnutella”, http://www.gnutella.com.
[4] “Skype”, http://www.skype.com/intl/en/.
[5] Zhang, R., Charlie Hu, Y.: Assisted Peer-to-Peer Search with Partial Indexing. IEEE Transactions on Parallel and Distributed Systems, Vol. 18, No. 8, pp.1146-1158 (2007).
[6] Sripanidkulchai, K., Maggs, B., Zhang, H.; Efficient Content Location Using Interest-Based Locality in Peer-to-Peer Systems. Proc. IEEE INFOCOM (2003).
[7] Stoica, I., Morris, R., Karger, D., Kaashoek, M. F., Balakrish-nan, H.: Chord: A scalable peer-to-peer lookup service for internet applications. In Proceedings of the ACM SIGCOMM ’01 Conference, San Diego California (2001).
[8] Maymounkov, P., Mazi`eres, D.: Kademlia: A Peer-to-peer Information System Based on the XOR Metric. In Proceedings of the 1st International Workshop on Peer-to-Peer Systems(IPTPS) (2002).
[9] Tajima K., Mizuuchi Y., Kitagawa M., Tanaka K.: "Cut as a Querying Unit for WWW, Netnews, and E-mail", Proceeding of 9th ACM Conference on Hypertext and Hypermedia, pp.235-244 (1998.6).
[10]Necip Fazil Ayan, Wen-Syan Li, Okan Kolak: "Automating Extraction of Logical Domains in a Web Site", In International Journal of Data and Knowledge Engineering, 43(2), Elsevier Science, pp.179-205 (2002).
[11]Mizuuchi Y., and Tajima K.: "Finding Context Paths for Web Pages", Proceeding of 10th ACM Conference on Hypertext and Hypermedia, pp.13-22 (1999).
[12]JIN Y., Matsuo Y., Ishizuka M.: "Extracting Social Networks with Weak Relationships from the World Wide Web", The IEICE transactions on information and systems (Japanese edetion) J91-D(3) pp.709-722 (2008).
[13]Murata T.: Discovery of Web Communities Based on the Co-occurrence of References, Transactions of the Japanese Society for Artificial Intelligence Vol.16 No.3 B, pp.316-323 (2001).
[14]L. N. Foner. Yenta: a Multi-Agent, Referral-Based Matchmaking System. In Proceedings of the 1st International Conference on Autonomous Agents (AGENTS ’97) (1997).
[15]“Overlay Weaver”, http://overlayweaver.sourceforge.net/.

Evidence preservation method using hysteresis signature for the portable terminal

Takashi Mishina*, Yoh Shiraishi** , and Osamu Takahashi**

*Graduate School of Systems Information Science, Future University Hakodate, Japan
**School of Systems Information Science, Future University Hakodate, Japan
{g2109042, siraisi, osamu}@fun.ac.jp

Abstract - The portable terminal is an important information terminal that contains a lot of individual information such as addresses, telephone numbers, e-mail addresses, telephone call history, etc. As corporate use of the portable terminal increases, it will be necessary to prove the cause of computer security incidents to decrease information leaks due to human factors. These factors include "ignorance" (unwittingly losing information), "fault" (performing the wrong operation or losing / stealing the portable terminal), and "intention" (intentionally selling information illegally).

We propose a technique to preserve information in the portable terminal to prove the terminal's behavior and how information has leaked. In the proposal, we consider two problems. One is the reliability of the evidence when we applied digital forensics to the portable terminal and the other is the few calculation resources of CPU and memory, etc.

Our proposal method solves these two problems by using the hysteresis signature and priority level of the evidence preservation.

Keywords: Digital forensics, Security, Hysteresis Signature.

1 INTRODUCTION

The portable terminal is an important information terminal that a lot of people have and is used for various purposes such as voice communication, sharing personal information in data communications, and internal and external access to the corporate system. Moreover, the network is temporarily constructed with the portable terminal, and the information is exchanged, for example, during natural disasters. Further developing portable terminal technology is expecting to produce even more applications.

In these situations, a lot of important individual information is used such as addresses, telephone numbers, e-mail addresses, telephone call history, operation logs, file information, etc. When individual information is used, information security measures are needed. Current information security measures have been designed chiefly to prevent invasion and operation from outside. However, information leaking from the inside to the outside is also a problem. Ninety percent or more of the information security breaches in which information leaks from the inside to the outside is attributed to three "human factors" [1]: "ignorance", meaning information is mistakenly thrown out; "fault", meaning operating incorrectly or losing / stealing the portable terminal; and "intention", meaning intentionally selling information illegally.

There are two methods for decreasing the breeches caused by the human factor. One is continuously educating the person. The other is giving the portable terminal physical measures and covering the person's mistake. The problem with education is the effect takes time and it costs a lot. Therefore, we apply digital forensics to the portable terminal, and pay attention to the method for decreasing the negative effects of the human factor.

Digital forensics [2] is a technology that gathers and maintains evidence in order to prove an illegal invasion, leak of information, etc. However, it has problems, such as sometimes not collecting information, its performance decreases and stops the system, detection is delayed, and evidence being destroyed during investigation [3]. Moreover, specific problems when digital forensics is applied to the portable terminal include "Enhancement of the telecommunication facility", "Data memory composed of flash memory", and unreliable evidence". [2]. However, the calculation resources are limited for the portable terminal, so it is difficult to secure reliable evidence using only the terminal and to always keep collecting all information.

Thus, in this paper, we aim to prove the information leaks caused by the human factor with a portable terminal and to prove how the terminal behaves. Also, we propose an evidence maintenance method in the portable terminal that considers how to reliably secure evidence, and the calculation resources such as CPU and the memory are limited.

2 RELATED RESEARCH

This chapter details an existing digital forensic method to the portable terminal and the hysteresis signature,

which is the signature technology to secure reliable evidence.

2.1 Forensic applied to portable terminal

2.1.1 Method of portable terminal memory

Willassen suggests two methods to investigate deleted information for the memory of the portable terminal [4]. This method listed seven pieces of information as the evidence.

・Images
・Sounds
・Multimedia messages
・WAP / web browser history
・Email
・Calendar items
・Contacts

In the first method, it is connected to the on-board flash memory tip directly and reads the content of the memory. In the second method, it uses the boundary scanning test, which is the inspection method of an IC tip called JTAG, and reads the content of the memory. Both methods need a physical connection to the portable terminal, but correct information cannot be found with a portable terminal alone.

2.1.2 Digital forensic method uses portable terminal

Kunii proposes a system that realizes digital forensics and manages files in a small-scale computing environment [5]. In this method, the PC files are distributing preserved. Each user proves the legitimacy of each other's file update histories and realizes digital forensics. The portable terminal is used as a signature generation device in this method. However, because information in the portable terminal is not collected, it is difficult to prove whatever has happened using the portable terminal.

2.2 Hysteresis signature

2.2.1 Outline of hysteresis signature

The hysteresis signature [6][7] is a technology to minimize the damage of leaks or estimate the signature generation key. This signature technology uses a normal e-signature method as a part of the component. Signature information is left for the signature history as a signature record every time the signature is generated. When signing, the latest signature record is taken partially from new signature information. As a result, the chain architecture is given between signature records. Specifically, the signature is generated by using the private key for data that unites the hash value of the latest signature record with the hash value of the document for the signature (Fig.1).

A part of the signature history is safely kept by the tamper resistant module. As a result, even if a past signature is forged, this can be detected by confirming whether it corresponds to the history.

When verifying it, in addition to the normal signature verification using the public key, the signature is verified in the history chain. In addition, it is possible to verify the signature corresponding to the signature record. Therefore, even if a person illegally tries to counterfeit the content of the document and the signature, he/she must counterfeit not only one document and the signature but also all the signature records because the chain architecture is between signatures. All this makes forgery of the signature difficult.

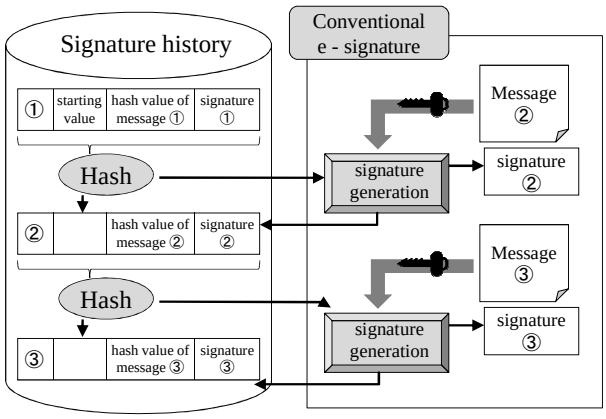


Fig.1 Hysteresis signature scheme

2.2.2 The signature history intersection

The signature history intersection is a chain method between one's signature history and others' signature histories used as the signature record. As a result, it is thought that the counterfeit becomes difficult as histories become longer because it needs the falsification of others' signature histories when a person illegally forges the signature (Fig.2).

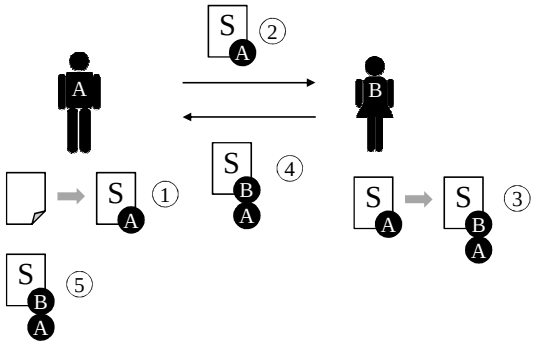


Fig.2 Signature history intercrossing

3 PROPOSED METHOD

This chapter describes requirements when digital forensics is applied to the portable terminal, priority level and collection frequency of information, and the flow of the proposed method.

3.1 Requirements

In digital forensics intended for the computer, usually data is maintained by the following process. To detect computer security incidents, to judged whether to collect data where it disappears by turning off the power supply.

Afterwards, power is turned off and data is maintained from the outside with some special equipment. The data that disappears when the computer is switched off are important, but collection activity can add changes to the data. However, there are various problems when applying the same method when digital forensics is applied to the portable terminal. Thus, the requirements when digital forensics is applied to the portable terminal are defined. They refer to the problem of the digital forensics that Ochi [3] enumerated and the problem when digital forensics was applied to the portable terminal that Tsujii et al. [2] enumerated. This paper determines requirements to prevent the following problems.

(1) Evidence is lost unexpectedly, and signs of an information leak cannot be acquired.

The portable terminal uses the flash memory for the data carrier and can delete all information by hardware reset. Therefore, signs of the relevant information leakage possibly cannot be acquired when evidence is maintained in the terminal.

(2) The data collection and the preservation of evidence degrade the performance of the portable terminal, and the system is stopped.

When data is collected, it is necessary to stop the system in the method of connecting the equipment from the outside. This method is unsuitable for the portable terminal, which is carried and used. When the performance of the system worsens, a lot of calculation resources are needed for the portable terminal in the method of frequently acquiring the bit stream image. The effect on the system must be reduced so that the contact in the emergency should not become impossible.

(3) When collecting evidence and maintaining it only with the portable terminal, evidence is not reliably secured.

If the private key leaks when the e-signature is put on data only for oneself, the counterfeit of the e-signature becomes possible. Moreover, evidence may be falsified by a malicious operation because the telecommunication facility of the portable terminal has been enhanced. Therefore, evidence must later be verified to make sure it is not falsified.

To solve these problems, we change information collection frequency by changing the information priority level and reduce the burden on the portable terminal. Additionally, we propose the method to secure reliable evidence by using the hysteresis signature.

3.2 Priority level and collection frequency

3.2.1 Priority level of information

The data of PC and the portable terminal exists in volatility and nonvolatile states. Nonvolatile data is the data (like filesystem stored in the hard disk drive and the flash memory) that continues after the computer is switched off. Volatility data indicates the disappearing data (like the present network connection of the system) when turning the computer off. Table 1 shows the list and the priority level of volatility and nonvolatile data of the portable terminal. Table 1 refers to the priority levels when data was collected that are generally recommended [8][9].

Table 1 Volatile data and nonvolatile data

		Volatile data	Nonvolatile data
Priority	1	Network connection	Config file
	2	Login session	Log file
	3	Running processes	Data file
	4	Opened files	Application file
	5	Network composition	Dump file
	6	Time of OS	Hibernation file
	7	Content of memory	

Priority levels are decided by considering the burden given to the portable terminal. "Content of memory" should acquire the memory image of the bit stream. Therefore, "content of memory" is the lowest priority level in volatility data. The bit stream image can generate a copy including the space domains of the original medium. However, a execution time longer and the burden on the terminal more than that for a logical backup that copies a file simply are needed. Because as many as [a 3rd to a 7^{th} of the nonvolatile data are needed to acquire the bit stream image, the priority level is low.

3.2.2 Change in collection frequency

With the priority levels that we showed in section 3.2.1, we collect and maintain evidence in three phases. Table 2 lists the collection range and frequency of the evidence.

Table 2 Collection frequency and collection range

Frequency	Range
High	Volatile data: priority 1 ~ 6
Medium	Volatile data: priority 7 Nonvolatile data: priority 1 ~ 2
Low	Nonvolatile data: priority 3 ~ 6

The high collection frequency acquires logical backup from volatility data priority level 1 to 6. The medium collection frequency acquires the image of the bit stream of RAM and logical backup from nonvolatile data 1 to 2. Finally, the low collection frequency acquires the image of the bit stream of ROM. By changing the collection frequency and range of information gathering, it is possible to prevent the system degrading the performance and being stopped by lowering the burden on the terminal. In addition, there is the difference in collection frequency, but the loss of evidence can be prevented to collect all information. As a result, we can solve problems (1) and (2) in section 3.1.

3.3 Algorithm of proposed method

The proposal method is composed of the portable terminal and the server. The portable terminal gathers and transmits evidence. The server secures maintenance and reliability of the evidence that has been sent. The server can be trusted enough like providers of digital certification services, and the access in the server must be severely limited. Figure 3 shows the flow of the proposal method.

3.3.1 Behavior of portable terminal

The portable terminal is processed as follows.

・Regular acquisition of evidence

・Generation of hysteresis signature

・Transmission of evidence and hysteresis signature

Evidence is regularly acquired by using the collection frequency and the range of the collection of evidence listed in Table 2. To minimize effects on the system, the program that gathers evidence is executed in the conserved region near a tamper resistant SD memory card. Moreover, to prove the completeness of evidence, the hash values with former data are compared. Collected evidence is stored in the SD memory card.

Next, the hysteresis signature is generated by using acquired evidence and signature history. Afterwards, evidence and the hysteresis signature are transmitted to the server, after which evidence is deleted and the hysteresis signature is preserved as the latest signature history. Information leaks are prevented by deleting evidence. When the hysteresis signature is received from the server, it is preserved as the latest signature history. The hysteresis signature is generated and the signature history preserved in separate processes in each range of the collection.

Evidence and the hysteresis signature are both transmitted an odd number of times, but only evidence is transmitted an even number times. By right, two or more terminals acquire data respectively, and the hysteresis signature is signed in the chain. However, in this process, only the portable terminal acquires data and a chain signature with the server is enabled. The reliable evidence can be secured by this process, and the problems described in section 3.1 can be solved.

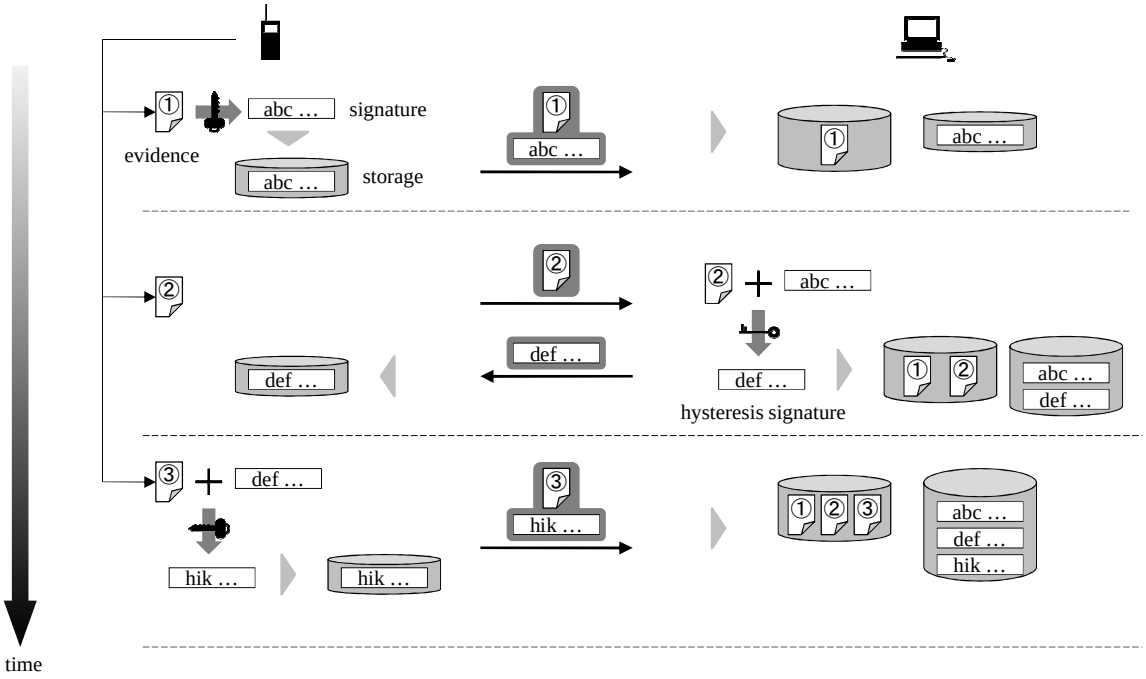


Fig.3 Flow of the proposed method

3.3.2 Behavior of server
The server is processed as follows.
・Reception of evidence and hysteresis signature sent from portable terminal
・Generation and transmission of hysteresis signature
All evidence and the hysteresis signatures sent from the portable terminal are preserved on the server side. The hysteresis signature is generated, and signature history preserved in separate processes in each range of the collection as well as the portable terminal side. The hysteresis signature is not generated and transmitted an odd number of times but an even number times (Fig.3). Because evidence is preserved only on the server side, evidence can be prevented from being falsified.

4 CONSIDERATIONS

This chapter shows the effects of the human factor with which the proposal method can deal and the results of the qualitative evaluation.

4.1 Incident that can be dealt with

The biggest cause of information leaks is "operational errors", in which information is leaked through wrongly sent e-mails. The second biggest cause is "Management mistake", in which important information is mistakenly thrown out with other information. In addition, there is losing or leaving behind mediums that include important information, theft, information illegally being removed, configuration errors, etc. Using the portable terminal by the proposal method enables these human factors to be proved.

(1) Ignorance
This includes management mistakes and configuration errors. When important information is thrown away by mistake, when, how and what was thrown away can be proven. At that time, running processes, time of OS, and data file are used. When the information leaks due to a setting mistake in the application, the cause can be proven. At that time, config. file and application file are used.
(2) Fault
This includes operational errors, loss or leaving behind and theft. What e-mail has been sent by mistake due to the operational error of the portable terminal can be proven. At that time, network connection, network composition, and log file are used. It can be proven if the portable terminal has been operated when the portable terminal has been lost or stolen. At that time, login session, running processes, and opened files are used. At this time, the proposed method should be able to prove that if someone had the portable terminal.
(3) Intention
This includes illegally removing information. The proposed method can prove what information is illegally taken out by using with the address and the telephone number, etc. At that time, data file and content of memory are used.

A lot of volatility data with high collection frequency is requested in these typical information leaks. It is thought that problems can be dealt with because a lot of volatility data and nonvolatile data are collected in the proposed method even when incidents other than those above occur. When the cause is proven, proof that keeps the temporal order is possible according to regular acquisition and the hysteresis signature of information.

4.2 Qualitative evaluation

The case where existing digital forensics of the computer is applied to the portable terminal is defined as "Non-apply" and is carried out as follows. The portable terminal is connected with a PC, and the image of the bit stream of RAM is acquired. Afterwards, the data of ROM is maintained as a bit stream image with the system of the portable terminal stopped. Table 3 shows the results of the qualitative evaluation by comparing the "Non-apply" and the proposal method.

Table 3 Qualitative evaluation result
(O: good , Δ: poor , X: no good)

	Non-apply	Proposed method
Miss acquisition	O	Δ
Performance degradation	X	O
Reliability	O	O

The evaluation item used the problem enumerated in paragraph 3.1.
(1) Miss acquisition of information
Non-apply collects and maintains the data of ROM and RAM as a bit stream image. Therefore, Non-apply come to be able to reduce miss acquisition of information by collecting and maintaining it when the moment security incidents happened.
The proposed method acquires the image of the bit stream of ROM and RAM. However, information frequently acquired is a physical data copy of volatility data. Therefore, nonvolatile data may be missed in some cases.
(2) The degradation of the system
In non-apply, there is no communication in an emergency because the portable terminal must be stopped.
In the proposed method, the collection frequency and the range of the collection are changed by using the priority level of information. Therefore, the burden on the portable terminal is small, so it is never stopped.
(3) Reliability of the evidence

To connect with a trustworthy PC and to gather evidence in non-apply, the reliability of the evidence is secured enough.

In the proposal method, collected evidence is preserved in the portable terminal once. However, it is thought that the reliability of evidence is secured enough so that hysteresis may be signed between the server and the portable terminal, and the servers preserve evidence for a long time.

As a result of the above-mentioned qualitative evaluation, when non-applying, the proposal method was inferior at preventing information being deleted. However, it was shown to prevent the performance of the system degrading, which is the most important thing in the portable terminal.

5 CONCLUSION

This paper proposed an evidence maintenance technique to change the collection frequency and the range of the collection by using the priority level of information when applying digital forensics in the portable terminal. Evidence that corresponds to a lot of information leaks can be gathered by frequently collecting important volatility data and reducing the burden on the portable terminal. Nonvolatile data is collected at low frequency. Therefore, this data can correspond to the data file and the application file that are requested as proof of an information leak. The signature using the portable terminal can be verified as evidence by using the hysteresis signature. Therefore, reliable evidence can be securely maintain. The utility of the proposed method was able to be confirmed by evaluating it qualitatively.

As a future task, it is necessary to do a fixed quantity evaluation of the proposed method. In this evaluation, the proposed method is mounted, and how the calculation resource of the system is consumed will be confirmed. Moreover, it is necessary to set up the optimum value of the collection frequency of evidence. It is necessary to collect a lot of evidence by using the optimum value while minimizing the load on the system.

REFERENCES

[1] Hisamichi Ohtani (Working Group Leader): Information Security Incident Survey Report, NPO Japan Network Security Association Security Incident Investigation Working Group (2008)

[2] Shigeo Tsujii (editorial supervisor): Digital forensics dictionary, Digital forensics society (2006)

[3] Takao Ochi, Takao Kojima, Masao Togawa, Yukio Itakura: The Proposal of Incident detection Method using the Hot Digital Forensic, Proc. IPSJ SIG Notes 2008, pp.267-272, Information Processing Society of Japan (2008)

[4] Svein Willassen: Forensic Analysis of Mobile Phone Internal Memory, Proc. International Federation for Information Processing (IFIP), Vol.104, pp.191-204 (2005)

[5] Yui Kunii, Ryuya Uda: A Proposal of A Distributed File Backup System for Digital Forensics Using Cellular Phone, Proc. Multimedia, Distributed, Cooperative, and Mobile Symposium (DICOMO 2009), pp.671-678 (2009)

[6] Seiichi Susaki, Tsutomu Matsumoto: Alibi Establishment for Electronic Signatures, Proc. IPSJ Journal, Vol.43, No.8, pp.2381-2394 (2002)

[7] Mitsuki Shinoda, Yusuke Ueda, Ryoichi Sasaki: Proposal and evaluation of hysteresis signature system with paper document, Proc. Technical report of IEICE. ISEC, Vol.103, pp.77-82

[8] Karen Kent, Suzanne Chevalier, Tim Grance, Hung Dang: Guide to Integrating Forensic Techniques into Incident Response, Proc. NIST Special Publication 800-86 (2006)

[9] D. Brezinski, T. Killalea: Guidelines for Evidence Collection and Archiving, RFC3227 (Best Current Practice) (2002)

Network Affordance in a Region by Active Reflection Nodes

Taizo Miyachi*, Yoshihito Mori*

* School of Information Science and Technology, Tokai University, Japan
miyachi@keyaki.cc.u-tokai.ac.jp

ABSTRACT

We discuss a basic framework of an ecological knowledge distribution system that supplies people useful knowledge in a region up to every corner in the society. We propose an effective network architecture based on Active Reflection Node (ARN) and Context Node that adapts knowledge for target contexts. ARN transmits useful knowledge to itself in order to evolve the knowledge in a target region and timely distribute it to people all over to every corner of the region.

Keywords: active reflection node, knowledge evolution, human action, Network affordance, ecological communication

1 ITRODUCTION

Major promising services in ubiquitous computing environments are (a) awareness service in a context and (b) solution service based on a context in regional environments. Useful knowledge in the region consists of general knowledge and deep knowledge. The volume of useful knowledge evolved in a context has been increasing since the knowledge is shown either by video or by animation. Frequent transmission of both deep knowledge and evolved knowledge is very useful for people up to every corner of the society in order to take actions up a notch in the flood of information.

In this paper we propose a "network affordance" that "active reflection node" and "context node" evolve a general knowledge into the useful "evolved knowledge" in a context in a region. We also discuss the benefit of the network affordance using network meta-information management.

2 USEFULE REGIONAL KNOWLEDGE

Solution service is one of the major targets in ubiquitous computing society in the u-Japan national policy instead of fast access to web and electronic commerce in the e-Japan national strategy. Useful knowledge in a region mainly consists of general knowledge, deep knowledge and scheduled knowledge based on the characteristic of local region such as geometric character, culture, nature, transport and communication infrastructure, etc.

Both context aware service and solution services based on observed information by various sensor networks is very important for timely solution service in a long time period.

Major five big changes in ubiquitous computing society should be considered for the service.

(i) **Lots kinds of targets in observation**. Targets in observation by sensor networks are extended not only to everything in the society but also to human actions.

(ii) **Users up to every corner in the society.** A service for each context in a target field should be provided for all people up to every corner in the society since all the people should know indispensable knowledge in case of natural disasters etc.

(iii) **A Long time period.** A time period for observation by sensor networks becomes longer such as a few hours and a year. Taking actions in an action list and solving problems need a certain time period.

(iv) **Large volume of knowledge**. The volume of information and knowledge has becoming large since videos and 3D videos will usually be used like those in "Youtube."

(v) **Easy terminal.** Users can easily access to knowledge service by easy terminals such as mobile phones, electronic book like iPad, Internet TVs, and digital signage etc.

2.1 Intelligent functions in six types of features of a network service system

Knowledge management middleware and effective network architecture for the context aware service should be enhanced corresponding to the changes in ubiquitous society. Network service should be fitted in the human society. Knowledge service should be merged with human actions in order to distribute the useful knowledge to people all over to every corner in the society.

Knowledge distribution and sharing system should have six types of intelligent functions and features.

(1) Six types of features in knowledge distribution

Knowledge distribution systems can be classified into six types. (a) Query from a user such as Web access, (b) Collaboration between TV and the Internet,

(c) Collaboration between Digital signage with geometric information and the Internet, (d) Environmental assistance (robot and advanced traffic signal etc.), (e) Foreign worker usage, and (f) Environmental usage (nature, science, social environment and international society). Seamless knowledge services are supplied the people among these six usages.

(2) Good knowledge distribution for regional culture

We should improve the cultural level in a region and make a variant city based on the local culture by good knowledge distribution. The true value of special product and the pleasure in an event should be transmitted to visitors in order to acquire a deep understanding and sympathy from them.

(3) Good knowledge distribution for human

A user can generally understand the first task but can not always fully understand the whole subsequent procedures when he/she starts a job since human can not remember many things more than the magic number [5]. Knowledge distribution system let a user be aware of new matters and allows the user to take a suitable actions corresponding to the change of situation in a certain time period based on user's back ground knowledge and the ability even in an emergency time. Original explanation should often repeatedly be distributed again at the right moment.

2.2 Success network in a regional context

People often do not know what to do in emergency time since they happen to encounter a new situation for the first time. "Success network" represents major success stories in an encountered context in a region with a list of actions in certain amount of time period based on user's background knowledge and ability. A success network consists of two layers (i) success context layer and (ii) success context creation layer.

(i) **Success context layer** (SC layer): Success network in the success context layer represents a list of actions with important knowledge in a context for a great success (See Fig. 1). Success context layer also allows a user to know the local knowledge and specific actions that are chosen for both a target region and a specific situation in order to solve the local problems. Success network has meta-data related to user's preferences and user's bias for the recognition of a target since the uses in a region usually have a different culture and different bias for the target.

(ii) **Success context creation layer** (SCC layer): Success creation layer in Success network has meta-data related to clues, bias, key information and time information in order to timely improve the chain of successes in SC layer. User's bias for a target issue changes both recognition of encountered context and decision of solution. The worst damage happened by simultaneous causes of user's wrong bias and sudden change in the environments. Switching between main target and second target corresponding to the change like stronger typhoon is managed in SCC layer.

(1) Distribution of knowledge in a region

The regions can be mainly classified into (a) geometric and cultural region and (b) region of a field.

Useful knowledge in an urgent context timely makes users aware of problems and avoids a variety of risks.

Example 1 in case (b): Abnormal weather.
Farmers should cope well with unusual heavy snow in April in Japan in spite that they could not suppose a lot of snow in spring. The repetition of warning for lots of snow in a long time period can make the farmers avoid the big damage.

Useful knowledge that is evolved in a context should not only make a user be aware of problems but also show how to solve them corresponding to the progress of the user. Regional knowledge should timely be transmitted for all users in a region on schedule in the decided condition. The useful knowledge in a long time period should also repeatedly be supplied to all users in a region at an appropriate time.

(2) Context management for the change of environments

Usefulness of problem solving is basically corresponding to a user's context that the user encounters. We propose a basic structure of "context."

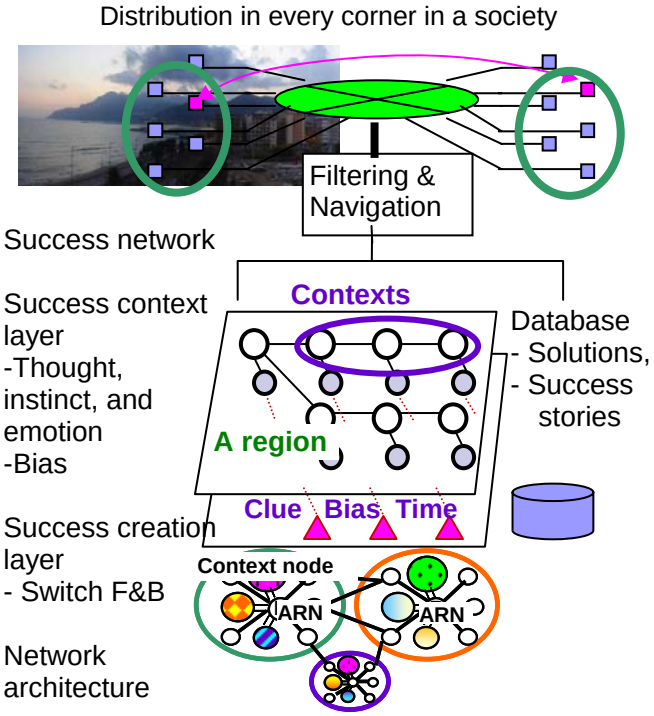


Fig. 1 An image of success network with contexts and active reflection nodes with context nodes

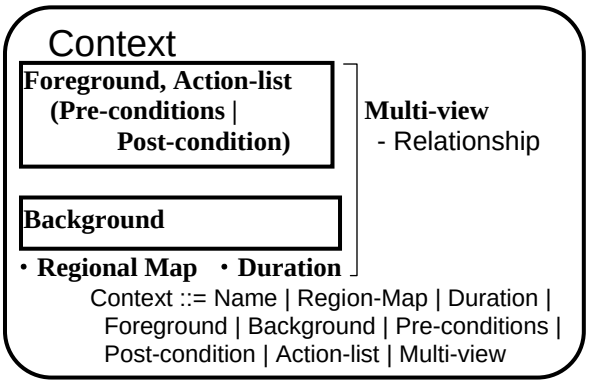
■**Definition of Context**

Context consists of nine attributes: Name, Region-map, Duration, Foreground, Background, Pre-conditions, Post-condition, Action-list, Multi-view. All attributes of context except for name are described in text, voice, image, animation and movie (See Fig. 2).

Name is not only a word but also a trigger of an image of context. Context has a target region and its map for knowledge distribution to all people in a target region. Duration means not only the exact time but also irregular occurrences in certain long time period. Examples of the duration are occurrence after five minutes, repetitions, by 3rd September 2010 and within 5 seconds. Context describes not only "foreground" but also "background." The foreground describes main targets. The background describes about the second targets. Background-conditions are very important in order to understand detailed context and to decide the order of actions. Background-conditions also describe preparation for actions in the context. Action-list describes a list of actions with Place, Position, Relationship, Time, Risk, Environment, and Potential. Time is very important for timely distribution of useful knowledge. Time describes not only the exact time but also both interval time between the former action and target action and a certain time period. Multi-view shows descriptions in multiple views with relationships in order to extend pleasure and to avoid risks. Avoiding risks should become in high priority in case that the relationship would mean indirect.

(3) Bottle neck of knowledge distribution

Lots of people can effectively access useful knowledge since success network timely shows it. This process produces plenty of popular knowledge in an area in a certain time period. Such knowledge is usually large volume of video like contents in "Youtube" since it is easy to understand up to deep knowledge. Therefore, such timely accesses would bring about massive accesses and cause a bottle neck of the Internet access.



⇒ **Guides for people in every corner of society**

Fig. 2 An image of structure of a context

3 ACTIVE REFLECTION NODE AND NETWORK AFFORDANCE

Timely problem solving for an encountered issue based on real time observation needs active assists by network system that manages reliable and effective communication in the Internet although a usual solution of a problem can be provided by a database system and its application systems. Typical problem solving in a focused region in a certain long time period should be assisted by the network system.

Useful videos enable most people in a target region to easily understand deep knowledge and to take effective actions in the background context. Concentration of accesses to a massive amount of knowledge needs effective management by the network system. Management of "network meta-information" such as reducing the bottle necks and saving energy of lots of server machines is very important for reliable solution services through the Internet.

3.1 Network architecture with active reflection node and context node

We propose new network architecture for effective distribution of the popular knowledge in a region considering meta-information [1, 2] such as "the volume of knowledge", "a long interval of knowledge access", and "local characteristic of access in a region." An example of the local characteristic of access in a region is the characteristic of a group of computers that access a same knowledge after an original knowledge is distributed for the registered people.

An "active reflection node (ARN)" sends a message to itself via "context nodes" (See Fig. 1 and Fig. 3). A "context node" is a kind of slow node to slowly transmit knowledge and has many kinds of views for target fields. Major roles of a context node are (1) estimating the risks based on information from many kinds of sensors for target regions, (2) gathering supplemental information and solutions based on the

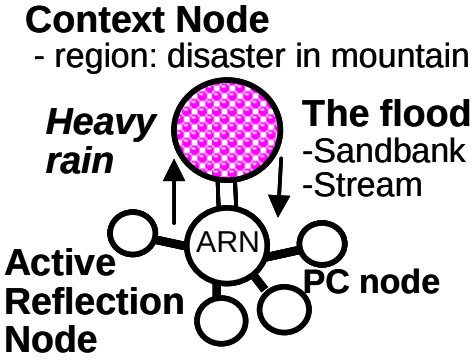


Fig. 3 An image of roles between active reflection node and context node

newest environments and (3) transmitting a merged solution with a schedule to the ARN and (4) forgetting useless knowledge. A context node can easily be added corresponding to a new focus.

ARN can quickly access the knowledge in context nodes. The ARN transmits the merged solution by a context node to all users in the region as a "network affordance" according to a defined schedule. This reduces the communication outside of the region and saves the energy consumed by the outside servers.

- Example 2. ARN transmit the deep knowledge scheduled in a long time period to all users who live in the beachside in a ria coastline even if the first seismic wave would be small. The deep knowledge is "Both second and third seismic wave are bigger than the first one. People should stay in the hill area."

The ARN architecture can save energy so as to find a set of the best solution in the region without plenty of Web accesses with large volume of knowledge out side of the region. Because all users in the region had better just access the nearest ARN in the region.

We measured the number of intermediate nodes that is passed through to the target web-site (Refer to Table 1). The number was same between in a busy time and usual day time. Google and Yahoo needed eleven intermediate nodes in a same Kanto area in Japan. Rakuten, Fukuoka-disaster-center in Japan and Youtube.com needed thirteen intermediate nodes. Okinawa-Island-disaster-center in Japan needed sixteen intermediate nodes. Foreign Web site except LA-times needed more sixteen intermediate nodes. Accesses to the nearest Web site in a same area needed between two and four intermediate nodes.

We suppose that ARN architecture can reduce two third of the node accesses. ARN architecture could reduce about eight node accesses for an access. For example about eight million server accesses for a million accesses by users could be reduced. Moreover,

Table 1 The number of intermediate nodes for a Web site access (TO: Time Out)

Target Web site	Nodes	Location
google.co.jp	11	Japan-H-Island
yahoo.co.jp	11	Japan-H-Island
rakuten.co.jp	13	Japan-H-Island
youtube.com	13	Japan-H-Island
fuku-bou.or.jp/	13	Japan-K-Island
okinawa.bosai.info	16	OkinawaIsland
latimes.com	8	US LA
Raileurope.co.uk	15+TO	UK
onlyinsanfrancisco.com	16+TO	US SF
australia.gov.au	16+TO	Australia
mlit.go.jp	17+TO	Malaysia
www.cc.u-tokai.ac.jp	4	Close area
pubweb.cc.u-tokai.ac.jp	4	Close area

eighty times long transmission for a million users could be reduced if the average length distance between intermediate nodes would be ten times longer than the distance between the nearest severs. Reducing such transmissions could obviously save lots of energy. The ARN architecture can also contribute to reduce bottlenecks out side of the region.

The ARN creates merged solutions among the context nodes by filtering the valuable knowledge in the target region as a result. Network affordance can be transmitted not by only data base systems but by network systems using network meta-information including geometric information.

The ARN allows the users to repeatedly execute the useful solution in a certain time period. The ARN architecture can also manage solutions in a long time period corresponding to the local characteristic in the region. ARN has four kinds of long time period transmission. (1) Sign oriented transmission, (2) Sensor oriented transmission, (3) Emergency oriented transmission, and (4) Due date oriented transmission. ARN can timely transmit a variety of delayed warning and solutions to all users in a region in order to avoid the damage of natural disaster etc.

3.2 Evolved knowledge in a region

Human can solve a problem with general knowledge in a region at the right moment. He/she can also reduce the risks of a wrong decision and misunderstanding with either deep knowledge or scheduled knowledge.

■ **Distribution of deep knowledge**

Human can not usually understand the whole knowledge when he/she acquires it. Not too much knowledge for his/her ability should be supplied step by step is important so as to deliver a deep knowledge such that meaning of the deep knowledge will be clarified after the process would be proceeded.

- Example 3 deep knowledge. Potato planting. Most people remind that they should cut a big potato into several pieces of small potatoes and plant them in the soil. However some people have forgotten the preparation of planting. People should burn some straws and make ash in order to spread it on the cut end of potatoes. This becomes sterilization and disease prevention. People should also cover a black plastic sheet on the planting bed and step on the edge of the sheet in order to multiply the soil on it from the side of the planting bed. The sheet keeps the planting bed warm and prevents the breeding of weeds.

3.3 Structuring knowledge for simplifying roles of human brain and extending them

People can build a basement for the success even in a hard situation by himself/herself after he/she experienced a success and a similar success in the success network. A structuring method of knowledge not to forget with only one listening is very useful for every one so as to easily understand knowledge [5]. This method includes an easy way to memory an important object by a pair of targets that a user remembers a target with a familiar object or a dangerous object.

Extracting and exploiting emotion of users, discovery of user's preferences, ranking of used knowledge by target users are very important for effective knowledge distribution. Knowledge for success is utilized for functions of human brain such as (a) recognition, (b) imagination, (c) perception / disillusion, (d) learning, (e) creation, (f) emotions, (g) reflection/introspection, (h) decision making, and (i) command for an action.

Transmitting user's favorite knowledge without stopping the functions in brain is important so as to assist human actions. The success network supplies a user with such knowledge in a target context and assists him/her to notch the great success. Evolved knowledge in a new context can show good specific actions for local characteristics in the target region although general knowledge shows a common way in a general context. Sharing such evolved knowledge navigates the user to the great success.

A good way of structuring flexible deep knowledge is very important for people all over to every corner in the ubiquitous computing society in order to adapt for an unusual condition. Experienced people and experts should build a highly evolved knowledge up a notch for the success in specified conditions in a region (See Fig. 4).

Carl Gustav Jung found that human naturally has collective unconsciousness that might be a base of symbiotic relationship between human [3, 4]. Evolved knowledge and shared consciousness can be created by people in a region based on reliable information accumulated by sensors and people in a region.

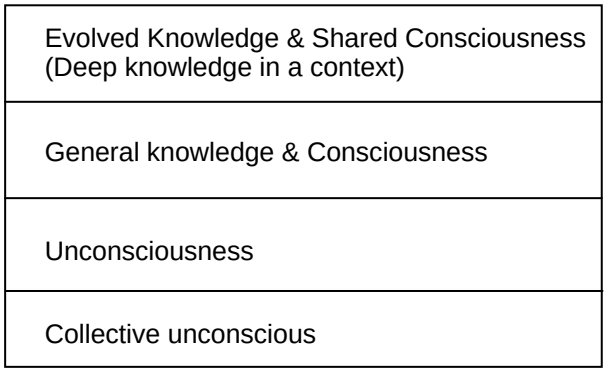


Fig. 4 Evolved Knowledge & Shared Consciousness

Positive unconsciousness and shared consciousness could work to assist human actions to the great success. Human should improve evolved knowledge and create shared consciousness corresponding to major contexts in the region in order to effectively develop a variant city and saves energy both in it and in the Internet.

3.4 Visualization of ARN and context nodes

Visualization of both the state of ARN and the kind of knowledge in a context node allows users to know the current state of knowledge distribution in a focused region and in a whole target area (See Fig. 5). The visualization also shows local emotion of users in a target region and enables marketing for a new product in a certain time period such as news release of it and a summer sale.

Manager of the network system can discover abnormal nodes and an abnormal region by a map of network. The manager can repair while observing the change of states.

4 NETWORK AFFORDANCE IN A REGION

Everyone in the ubiquitous society will be able to easily utilize both general knowledge and evolved knowledge in a corresponding context by the success network using the easy terminal any where any time.

The knowledge becomes more valuable in case that the other people do not help a victim in serious situation since the other people can not know who should be helped before the heavy damage happens.

4.1 Network Affordance from natural environments

Evolved knowledge can be utilized even in case of nature disaster such as flood by sudden heavy rain, hailstone in spring, and typhoon. For example, a victim

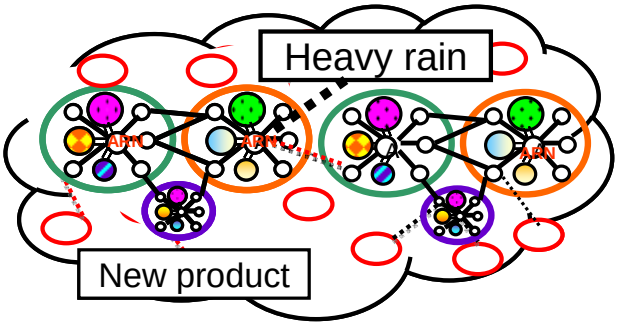


Fig.5 An image of display of colored context nodes and ARN for a heavy rain by a seasonal rain

can timely foresee the risk of a natural disaster by sensor networks. The victims and experts create an evolved knowledge and enable people all over to every corner of a target region to check the dangerous spots and to know how to avoid the risk using the evolved knowledge in a success network in a dangerous context.

Example 4 Flood by sudden heavy rain

People who enjoy camping in a sand bank in a river can know the sudden flood even at night by success network. Success network can show action plans in the evolved knowledge that is a result of analysis of both sensor data in a west area accumulated in several years and geometric characteristic of the river. A mobile terminal could timely receive an emergency mail from the Web site of success network and rings a bell in order to let the user know the danger at the right time. People can know the danger of flood even if there is little rain in the sand bank.

Observation systems could play a same role in the mobile terminal using success network since they send the data related to the danger to the context nodes. The system can tell about the danger and explain action plan to avoid the damage by loud speakers surrounding the people. The explanations by both the observation system and the mobile terminal produce an affordance from natural environments based on the sensor network etc.

Example 5 Disease of a baby at night

A mother can be assisted in case of disease of baby at night by general knowledge with a photo that shows baby's mucus. She can know that the baby is out of danger since the photo shows her an example of safe case of baby's mucus. She does not have to bring her baby into a danger room in a hospital where lots of disease-causing bacteria lives.

She can also take the best action for the disease of the baby utilizing action plans with effective route navigation by the success network that reflects the location of emergency hospitals near her house and the availability of them.

4.2 Network Affordance from city environments

Environment assistances such as robots and advanced traffic signals keeps observing human actions and mobility and timely navigates to the success by transmitting knowledge as an affordance from surroundings in a city [6]. Parametric speaker without sound pollution is suitable for the guide in a city that hates the sound damage since knowledge can be provided from 360 degree surroundings by hearing voice guides while moving, taking an action, and watching other objects. (See Fig. 6)

A pedestrian can hear useful explanation with accurate direction and ensure the direction to go to the destination. Human can simultaneously listen to multiple voice guides reproduced close to ears by the parametric speakers. The guide would be transmitted as an affordance translated from sensor information such as crosswalking technique in real time.

There are many kinds of information terminals in the city. Digital signage systems in the cities also show knowledge and CM in large displays in order to enjoy the city life. A digital signage in a super market explains how to cock delicious food utilizing foods left in the refrigerator in order to sell their products. Information displays in a station guide visitors to sightseeing spots, restaurants and hotels. Train vision in trains also shows the current state of transport and allows users to download useful knowledge by wireless communication.

5 CONCLUSION

We propose "success network" based on contexts in order to provide suitable actions for the great success in a region in the big change of environment.

"Active reflection node" architecture in order to effectively transmit evolved knowledge to people all over to every corner in the society in a certain long time period is also discussed. This architecture can reduce both the number of servers and the total amount of transmit in the Internet in order to save energy in the Internet.

We also discuss "network affordance" in the context in a region by evolved knowledge that can adapt for both the change of environments and the ability of human. We expect that a set of evolved knowledge would create a shared consciousness and enable both residents and visitors to enjoy the life in a variant region.

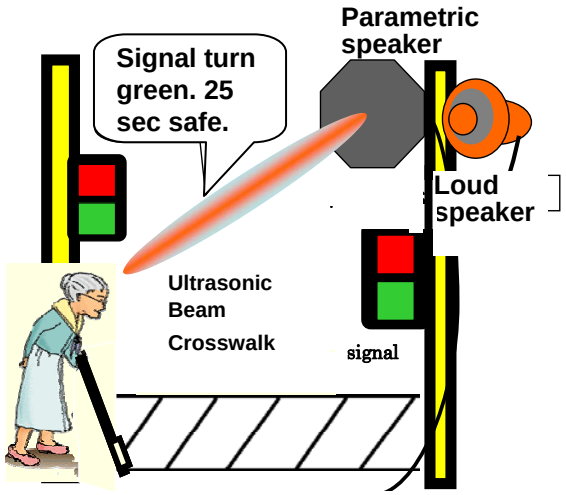


Fig. 6 Transmitting affordance by parametric speaker

REFERENCE

[1] A. Villazon, A Reflective Active Network Node. Springer, LNCS, Vol. 1942/2000. (2000)
[2] D. Hughes, et al., A Framework for Developing Reflective and Dynamic P2P Networks (RaDP2P), Proc. of the Fourth International Conference on Peer-to-Peer Computing (P2P'04). (2004)
[3]R. Nakatsu, Future of Communications Empowered by Technologies, Ohm-sha Publishing Co. (2010)
[4] Jung, Carl. Archetypes and the Collective Unconscious.(1959)
[5]T. Miyachi et al., Structuring Spatial Knowledge and Fail-Safe Expression for Directive Voice Navigation, Springer, LNAI 5179,pp.664-672. (2008)
[6] T. Miyachi, et al., A Quiet Navigation for Safe Crosswalk by Ultrasonic Beams, Springer, LNAI 4694, pp.1049-1057. (2007)

An Experiment for an Interactive Internet Live Broadcasting System with a High-Quality Snapshot Function

Yoshia Saito* and Yuko Murayama*

* Faculty of Software and Information Science, Iwate Prefectural University, Japan
{y-saito, murayama}@iwate-pu.ac.jp

Abstract –The video quality on the Internet does not come up to TV quality yet. Although audience often ask for high-quality videos, it is difficult to improve the quality because of its huge video traffic. In this paper, we propose an Internet broadcast system with a high-quality snapshot function to improve user experience. While the proposed system delivers low-quality video to audience, it provides a high-quality snapshot function which enables the audience to take a snapshot of a desired and favorite scene anytime. To assess effectiveness of the snapshot function, we designed and implemented a prototype system. This paper reports results of an experiment in a commencement ceremony of our university using the prototype system.

Keywords: Internet Broadcasting, Interactive TV

1 INTRODUCTION

With the spread of broadband Internet connections, Internet users can watch video contents all over the web in recent years. Video players are integrated into most web browsers and the Internet users can watch videos easily on their web browsers without any installation of proprietary client software. These video platforms such as Adobe Flash and Microsoft Silverlight rapidly popularize video viewing on the web. Tens of millions of videos are shared in video sharing services like YouTube [1] and enormous numbers of on-demand videos are watched by Internet users in various countries. Furthermore, live video broadcasting by Internet users also gets popular nowadays. A lot of Internet users broadcast their original live video contents using PCs and web cameras. It is expected that these video sharing and broadcasting services would become widely used much further and network traffic of the videos would grow more in the next couple of years.

The huge video traffic, however, causes a problem of communications expenses. Although most video sharing and broadcasting services run on income from advertisements on their websites, it is difficult to make profits because of its expenses more than its advertising income [2]. Moreover, current online video services distribute videos with a few hundred kilobits per second (kbps). The video quality on the Internet does not come up to TV quality yet. While the video services should provide more high-quality videos, it is not easy to improve the video quality for the above reason.
There are two solutions to increase user experience; one is to improve video quality upgrading video resolution and bit rate, and the other way is to enhance added value of the video contents introducing interactive features and so on. Under present circumstances, the later is a realistic solution. Interactive television (iTV) [3-6] is a research area which provides interactive features to video contents in order to improve user experience. Ustream [7] and Justin.tv [8] which are typical services for live video broadcasting apply the iTV technologies to their system. In these services, live video viewers can communicate with broadcasters and other viewers using chat and social communication tools (e.g. Twitter) watching live video contents. These interactive functionalities are attracting the attention of many Internet users despite low resolution and bit rate of the live videos.
Meanwhile, we have been trying and conducting several experiments with Internet broadcasting in graduation ceremony of our university [9, 10]. In these experiments, we had to deliver live video over the Internet with a few hundred kbps because we only have 100 Mbps connections between our university and the Internet. Therefore, it was difficult to satisfy parents who could not attend the ceremony and would like to watch their children's proud moment because of its low-quality video. We needed to add something extra to our live broadcasting system introducing special functionalities to improve user experience without increasing network traffic.

In this paper, we propose an interactive internet live broadcasting system called *Photographable TV* which provides a high-quality snapshot function so that audience can take high-quality pictures of favorite scenes for their memories at any time watching live video. In case of graduation ceremony, parents of graduates can take ceremonial pictures remotely as if they were attending the ceremony. The pictures can be saved to local disks for their personal memory albums. Since the data size of still pictures is far small than that of video, the proposed system can improve user experience without increasing network traffic. To study the effectiveness of the high-quality snapshot function, we design and implement a prototype system. We also conduct an experiment in our graduation ceremony to evaluate how to use our system by audience and find issues.

The paper is organized as follows. In the next section, we describe related work discussing originality of our proposed system. Section 3 introduces the model of the Photographable TV system. Section 4 presents the design of the prototype system of the Photographable TV, its system architecture and user interface. Section 5 evaluates the prototype system and reports the experiment results. Section 6 gives some conclusions and our future work.

2 RELATED WORK

To increase video quality for improvement of the user experiment, there are many researches. Typical one of them is IP multicast [11, 12] In the IP multicast, a sender transmits a single data stream to the receivers. Since the

routers on the path to the receivers replicate the data stream so that multiple receivers can receive it if required, it can deliver high-quality videos without increasing network traffic on the sender. However, it is not easy to use over the Internet because all routers on the path to receivers must support IP multicast. On the other hand, unicast is widely used in the Internet although it delivers multiple same copies to each receiver. This is because the unicast does not need special functionalities of the routers and can be used in any different network environments. In this regard, our proposed system uses unicast for the video delivery over the Internet.

Nowadays the P2P technologies are popularly used in the Internet to distribute network traffic over the Internet [13-16]. In these researches, hosts built an overlay network on the Internet by the P2P technologies and forward the received data stream to the other hosts like overlay multicast so that it can avoid concentration of the network traffic on a sender. However, the P2P technologies often require proprietary client software. The installation of the software prevents Internet users from casually watching videos with these P2P technologies. Moreover, the P2P software is often prohibited to use in a particular environment such as office network and university network because it goes through firewalls ignoring its network architecture. For these reason, it is difficult to widely use P2P software and enhance video quality over the Internet without increasing network traffic on a sender.

For similar ideas to our high-quality snapshot function, there are several studies in educational system. Ichimura proposes Chalk Talks [17] which is a remote lecture system with high-quality pictures. The Chalk Talks uses a HDV camera to provide a lecture with high-resolution. Since the high-resolution video consumes network resources, it compresses the video for the Internet broadcasting. The Chalk Talks also provides high-quality pictures at fixed intervals to clients so that the students can watch the white board clearly. While the Chalk Talks provides high-quality pictures at fixed intervals, the Photographable TV provides high-quality pictures when audience requests. In addition, the Photographable TV aims to make a personal memory album so that audience can remember the live broadcasting for improvement of the user experience although the Chalk Talks aims to improve readability of the white board.

One of video sharing services, PANDORA.TV [18] provides a snapshot function to the viewers. In the service, there is an image capturing button on the video player and still pictures of favorite scenes can be captured in JPEG format watching videos. However, it does not provide high-quality pictures because resolution and quality of the pictures are same as that of videos. Our system offers high-quality pictures to the audience more than video quality.

3 PHOTOGRAPHABLE TV

Photographable TV is an interactive broadcasting system for Internet live video streaming for the purpose of improvement of user experience without increasing network traffic by a high-quality snapshot function to enjoy high-quality pictures of favorite scenes for a personal memory album. The high-quality pictures do not increase network traffic so much if audience sends requests in moderate intervals. Even if the video quality is not so high, the high-quality picture would improve user experience and activate social communication on the web.

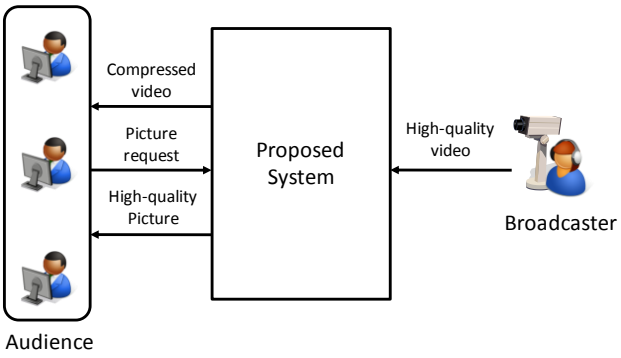


Figure 1. System model

Figure 1 shows the system model of the Photographable TV. This model consists of a broadcaster, its audience and the proposed system. Firstly, the broadcaster sends a high-quality video source to the system. The system receives and encodes the high-quality video storing the original source. The audience receives the compressed video from the system in real time over the Internet and also can send a picture request to the system anytime watching the video. When the system receives the picture request, a high-quality picture is made from the stored original video source and sent to the audience. The audience can see and save the high-quality picture.

There are several issues to realize the Photographable TV. The Photographable TV requires encode functions for video and pictures. Since quality of pictures is equivalent of the original video quality, the original video should be uncompressed and high-resolution so that high-quality pictures can be made from it. However, it is difficult to send the original video over the Internet because the data size of the uncompressed and high-resolution video is too large. This is an issue. The encode functions must be near the broadcaster side not to across the Internet. Besides video encoding, the broadcaster's PC has to extract a frame from the video and encode the frame to make a still picture. It is expected to consume CPU resource of the PC and we should take care of its load. Another issue is frequency of the high-quality picture requests from audience. The proposed system is available in accordance with an idea that picture traffic is much less than video traffic. If the audience frequently requests high-quality pictures, the picture traffic would be considerable amount. We have to study how many times the audience requests the high-quality pictures and control the picture traffic not to exceed network capacity.

4 PROTOTYPE SYSTEM

We developed a prototype of the Photographable TV to conduct experiments for its evaluation. In this section, we describe the design and implementation of the prototype system.

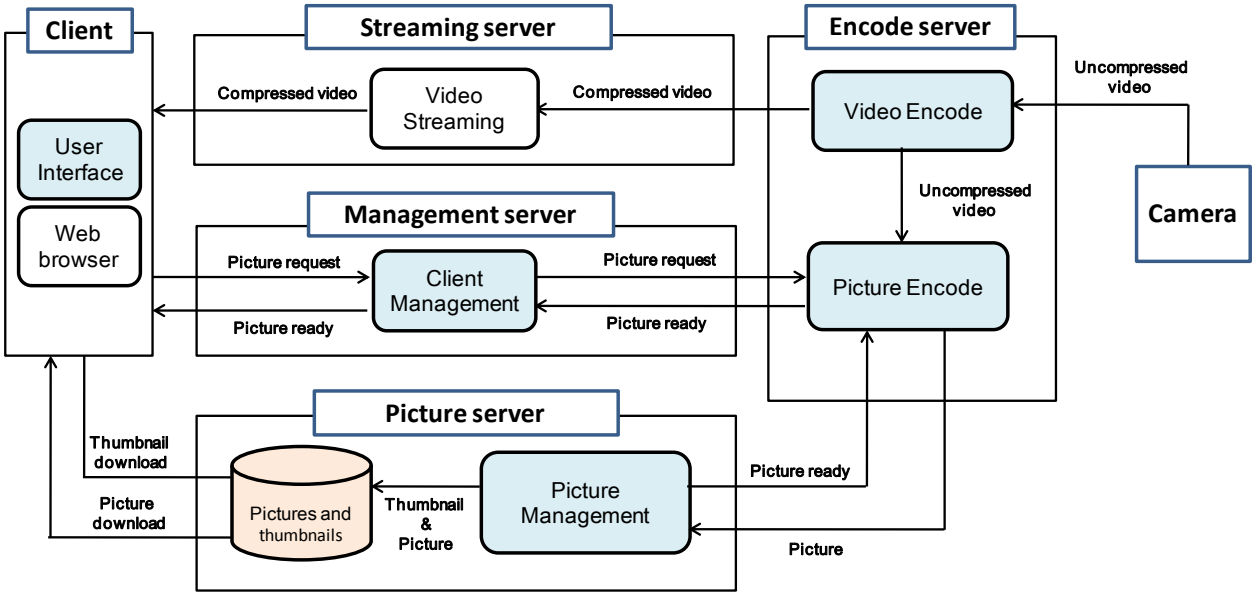


Figure 2. System architecture

Table 1. Server specifications

Picture server	
OS	Windows Server 2003
CPU	Intel Xeon 2.00 GHz
Memory	1024 MB
Web server	Apache 2.2.14
PHP	PHP 5.3.1

Video streaming server and management server	
OS	Windows Server 2003
CPU	Intel Xeon 2.00 GHz
Memory	1024 MB
Streaming server	Adobe Flash Media Server
JAVA	JDK 5.0

Encode server	
OS	Windows Vista
CPU	Intel Core2 Duo P8700 12.53GHz
Memory	4096 MB
Flash	Flex SDK 3.0
Video camera	Panasonic NV-GS320-S

4.1 System Design

Figure 2 shows the system architecture of the prototype system. The system consists of four servers; an encode server, a streaming server, a management server and a picture server. The encode server has two functions. One is a video encode function and the other is a picture encode function. The video encode function receives an uncompressed video source from a camera and compresses the video for broadcasting. The uncompressed video source is also passed to the picture encode function. The picture encode function stores the uncompressed video so that high-quality picture could be made from the source.

The compressed video is sent to a video streaming function on the streaming server. The streaming server sends the video to each client by unicast when requested. The audience can send a picture request to a client management function on a management server watching the video through a user interface on the browser of the client when they would like to take pictures of specified scenes. The client management function keeps client IDs and forwards the picture requests with their client IDs to the encode server. When the picture encode function receives a picture request from the management server, it encodes high-quality picture from the uncompressed video. The encoded picture is sent to a picture management function on a picture server. A thumbnail is made from the picture and they are stored in a database on the server. After that, a picture ready message is sent to the picture encode function and it is forwarded to the client management function with the client ID and location information of the thumbnail and the picture. The client management function forwards location of the pictures based on the client ID. The client only downloads the thumbnail from the picture server to save network resource in case the audience does not like the shot. After the audience confirmed the thumbnail, the client downloads the picture from the picture server and displays it on the user interface. The audience can save the high-quality picture to the local disk on the client to enjoy the pictures after the broadcasting.

4.2 Implementation

We implemented the prototype system based on the system design. The video/picture encode function on the encode server and the user interface on the client are implemented based on Adobe Flash written in ActionScript 3.0 for ease of video delivery and viewing over the Internet. The client management function is implemented by JAVA because it works on various environments. The picture management function is implemented by PHP script on a web server so that it is easy to upload pictures on the web server and make thumbnails of the pictures. We used Adobe Flash Media Server for the video streaming on the streaming server. The specifications of these servers are shown in

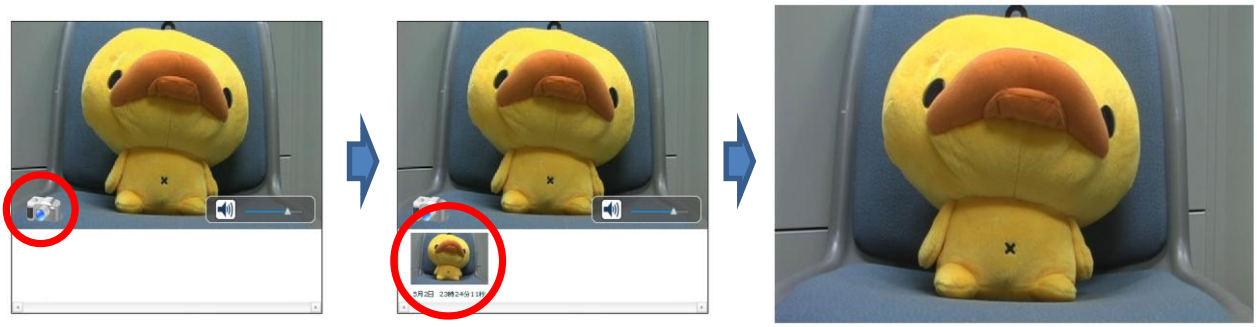


Figure 3. User interface for clients

Table 1. We used same PC for the video streaming server and the management server.

Video Encode

The video encode function receives raw video data from the camera with resolution 720x480. The raw video is resized and compressed by Flash framework to broadcast it in real time. Then, the compressed video data is sent to the streaming server in few hundred kbps. A still picture is captured in BMP format and passed to the picture encode function.

Picture Encode

Since the BMP file is not compressed, the file size becomes about 1 MB for a 720x480 still picture. It is too large to upload the file to the picture management server over the Internet. Therefore, we compressed the BMP data in JPEG format. After the compression, the files size will be from 30 KB to 50 KB and the network traffic between the encode server and the picture management server can be reduced.

When the BMP data is compressed in JPEG format, it increases CPU load of the encoder server. If the clients frequently send picture requests to the encoder server, the picture generation would be aborted. To make matters worse, the frequent requests would cause huge network traffic between the encode server and the picture server even if the data size of the JPEG files is small. Therefore, we introduced periodic picture buffering scheme into the picture encode function. The picture encode function stores BMP data on the memory at fix intervals. In this implementation, we set the interval to 500 msec taking into account the server load. When a picture request is arrived, the encode function searches latest picture from the arrival time minus video buffering time of the client on the memory. The BMP data is encoded in case it was not previously encoded. If the BMP data has been already encoded, it does not process the picture encode and returns only the picture URL to the client. The buffering scheme can reduce the server load and network traffic between the encode server and the picture server.

Picture Management

The compressed JPEG picture is sent to the picture server. The picture management function receives the picture. At the same time, a thumbnail of the picture with resolution 120x90 is made from the picture. The data size of the thumbnail is a few Kbytes. The high-quality picture and its thumbnail are saved in a public directory on the local web server. The client receives URL addresses of the picture and thumbnail. Although there is no user authentication to see the pictures in the prototype system, access control technologies should be introduced so that the pictures can be accessed by audience who owns them.

Client Management

We implemented four functions for the client management; 1) client ID management, 2) connection management, 3) picture request forwarding and 4) logging. The function of client ID management generates and keeps client IDs for each client which accesses to the management server. The client ID is a unique 22 characters and sent to the client when connected to the management server for first time. The client keeps the unique ID as a cookie on the web browser and can use the same ID thereafter. The ID is used for the picture request forwarding and system logging to identify the clients. The connection management function associates the client ID with its connection. The function of picture request forwarding notifies the encode server of a picture request with a client ID and replies the result to the client based on the ID. The logs maintain connected and disconnected time and the picture requests with their IDs.

User Interface for Clients

Figure 3 shows the user interface for clients. The compressed video is shown on the upper portion of the interface. The camera icon is a snapshot button to send a picture request to the management server. When the camera icon is clicked, it is not available before its response arrival. After completion of a picture request, a small snapshot picture is added to the thumbnail list by downloading from the notified URL address. A high-quality picture is displayed with resolution 720x480 on the other browser

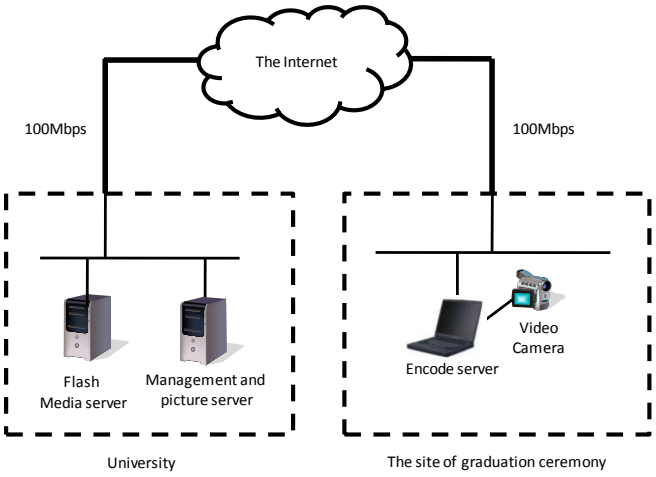


Figure 4. Network configuration

window when the thumbnail is clicked. The high-quality picture can be saved to the local disk of the client by an image saving function of the web browser.

5 EXPERIMENT

We conducted an experiment in our graduation ceremony with the prototype system in order to evaluate how to use our system by audience and find issues.

5.1 Methodology

Figure 4 shows the network configuration in the experiment. A flash media server and a management/picture server were placed in the University. These servers connected to the Internet at 100 Mbps. The graduation ceremony was held in the other place and we prepared 100 Mbps connections for the venue. An encode server and a digital video camera were employed. The bit-rate of the video streaming was 200 kbps. Any Internet users could watch the video streaming in real-time on our website for the broadcasting. We recorded number of the viewers, CPU load of the encode server and a log of photo requests on the picture/encode server.

5.2 Results

We analyzed the results to know how many people used and its scalability. Note that we excluded broadcasters and researchers related to the experiment from these results.

At first, we counted the number of the viewers through the broadcastings to study how many/long people used our system. Figure 5 shows the number of the viewers per second from 13:00 to 16:00. The total number of unique viewers is 148. The maximum and average numbers of the viewers are 43 and 33 respectively. As a whole, the prototype system kept the number of viewers throughout the broadcast.

We also analyzed CPU load of the encode server and number of the photo requests per second to study its scalability. Figure 6 shows the results from 13:00 to 16:00. From the graph, the CPU load was around 40%. Although the load momentarily marked around 80%, the encode PC

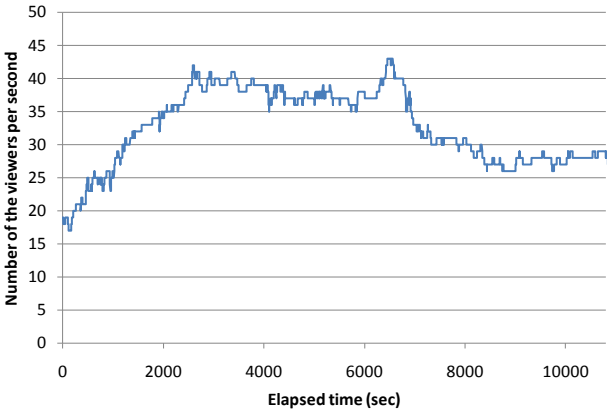


Figure 5. Number of the viewers per second

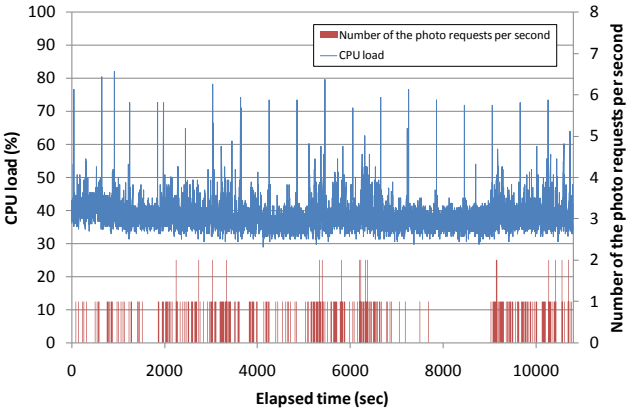


Figure 6. CPU load of the encode server and number of the photo requests per second

remained power throughout the experiment. While the number of the photo requests per second constantly occurred, the prototype system could provide the snapshot function. Since the total number of the photo requests was 423, the viewers used the snapshot function frequently.

From the experiment, we found the prototype system could be used by several tens of viewers at least and load of the encode server was suppressed by the periodic picture buffering scheme. The prototype system could provide the snapshot function for small-scale live broadcasting.

6 CONCLUSION

In this paper, we proposed an Internet broadcast system with a high-quality snapshot function toward improvement of user experience. The proposed system delivers low-quality video to audiences. Meanwhile, it provides a high-quality snapshot function which enables the audiences to take a picture of a desired and favorite scene anytime. We designed and implemented a prototype system and evaluated the system in our graduation ceremony. From the result, prototype system worked stably throughout the experiment even if more than 40 users watched the broadcasting simultaneously and the snapshot function was used 423 times. We confirmed the prototype system could provide the snapshot function for small-scale live broadcasting.

As future work, we will study server load when audience increases and its scalability. We will also conduct experiments at various events to evaluate the Photographable TV further in practical situations.

REFERENCES

[1] YouTube, http://www.youtube.com/

[2] Huang, C., Li, J. and Ross, W. K., "Can Internet Video-on-Demand be Profitable", SIGCOMM'07, pp. 133-144, 2007.

[3] Jensen, J., "Interactive Television - A Brief Media History", In Proc. of EuroITV2008, pp. 1-10, 2008.

[4] Cesar, P. and Chorianopoulos, K., "Interactivity and user participation in the television lifecycle: creating, sharing, and controlling content", In Proc. of UXTV2008, pp. 125-128, 2008.

[5] Ursu, M. et al, "Interactive TV narratives: Opportunities, progress, and challenges", ACM TOMCCAP, Vol. 4, Issue 4, 2008.

[6] Metcalf, C. et al., "Examining presence and lightweight messaging in a social television experience", ACM TOMCCAP, Vol. 4, Issue 4, 2008.

[7] Ustream, http://www.ustream.tv/

[8] Justin.tv, http://www.justin.tv/

[9] Takamizawa, N. and Murayama, Y., "Use of Telepresence in Informal Broadcasting over the Internet, In Proc. of the ACM SIGMM 2004 Workshop on Effective Telepresence", pp. 12-15, 2004.

[10] Saito Y. and Murayama, Y., "A Proposal of an Interactive Broadcasting System for Audience-driven Live TV on the Internet", Journal of Information Processing, Vol. 18 pp. 26-37, 2010.

[11] Deering, S.: Host Extensions for IP Multicasting, RFC 112, 1989

[12] Eriksson, H., "MBONE: The Multicast Backbone", Communications of the ACM, Vol. 37, No. 8, pp. 54—60, 1994.

[13] Tran, D., Hua, K. and Do, T., "Zigzag: an efficient peer-to-peer scheme for media streaming", IEEE INFOCOM 2003, pp. 1283-1292, 2003.

[14] Guo, M. and Ammar, M.H., "Scalable live video streaming to cooperative clients using time shifting and video patching", IEEE INFOCOM 2004, pp. 15001-1511, 2004.

[15] Cui, Y., Li, B., and Nahrstedt, K., "oStream: Asynchronous streaming multicast in application layer overlay networks", IEEE JSAC, Special Issue on Recent Advances in Service Overlays, Vol. 22(1), pp.91-106, 2004.

[16] Qu, C., Li, B. and Zhao, S., "Exploring large-scale peer-to-peer live streaming topologies, Transactions on Multimedia Computing", Communications, and Applications (TOMCCAP), Vol. 4, Issue 3, 2008.

[17] Ichimura, S., "Delivering Chalk Talks on the Internet", 40th Hawaii International Conference on System Sciences (HICSS), Collaboration Systems Track, IEEE CS Press, p.4-11 , 2007.

[18] PANDORA.TV: http://www.pandora.tv

An Implementation of WSN Data Management System on P2P Network

Nobuhiko MATSUURA†, Hiroshi MINENO‡, Norio SHIRATORI††, and Tadanori MIZUNO*

†Graduate School of Informatics, Shizuoka University, Japan
‡Faculty of Informatics, Shizuoka University, Japan
††Research Institute of Electrical Communication, Tohoku University, Japan
* Graduate School of Science and Technology, Shizuoka University, Japan
{matsuura, mineno, mizuno}@mizulab.net

Abstract - A context-aware service that uses sensing data has attracted attention, along with the development of wireless technology and sensor technology. To provide these services, the sensing data sharing system in P2P networks needs to cope with a vast amount of data. However, existing algorithms do not respond to varying the number of sensing data types. In addition, most existing algorithms cannot execute reverse key resolutions because their search algorithms need to include specific data as the key in the query. To address these issues, we propose a multi-dimensional range search algorithm in P2P networks that uses a B+tree for an efficient search with an arbitrary number of sensing data types.

Keywords: P2P, B+tree, range search, multi-dimensional search, wireless sensor network

1 Introduction

Peer-to-Peer (P2P) networks are emerging as a new paradigm for structuring large-scale distributed systems. In these systems, resources are associated with keys, and each peer is responsible for a subset of the key to guarantee the performance of scalability, fault-tolerance, and robustness. P2P have been developed to have a more suitable and practical design for applications such as those in [1], [2] because of their potential.

One of the systems that is suitable for using P2P is a sensing data sharing system, such as [3]. A context-aware service has attracted attention, along with the development of wireless technology and sensor technology. This service can offer a remarkable transformation that considers user location and conditions using sensor data. To provide these services, the system needs to manage data from wireless sensor networks (WSNs). P2P is thought to be the answer to cope with the vast amount of data from WSNs because of their potential.

However, WSNs have some dynamic properties such as varying the data values, the total number of data, and the number of data property(such as Temperature and Humidity). In other words, P2P must deal with these properties to perform as general middle-ware for a WSN management system because what it takes to provide service differs depending on the service. In particular, we are sure that considering the varying numbers of data property can improve the search performance because the sensor types will likely rapidly increase as the fundamental technology is developed. But other works in P2P networks could not deal with the varying number of data property. Their algorithms needed to include specific data, such as location information, as a key in the query. Therefore, they could not execute a reverse resolution of keys and could only use a limited number from a vast number of sensor types in the query condition. To solve the problem, we need the Multi-dimensional algorithm without the special property on the data store "Multi-dimensional" denotes the tabular form in data storage, and the query conditions number dynamically increases or decreases on demand. In this paper, we extend our previous work [4] and show simulation results to demonstrate the potential of our algorithm as WSN menagement system.

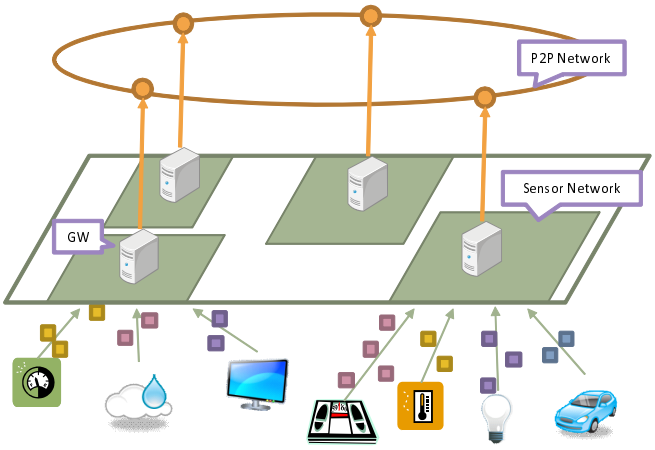


Figure 1: Sensing Data Sharing System

The rest of the paper is organized as follows: Section 2 discusses related work; Section 3 presents the our previous work and the problem; Section 4 show the experimental results and discussion; finally, section 5 concludes the paper.

2 Related Work

Much work has been done on sensing data sharing systems using P2P. These architectures arrange Gateway (GW) in WSNs to manage data transfers at each WSN and to construct structuring P2P networks automatically at each GW, like Fig. 1. In general, structuring P2P networks are constructed by using DHT, but they have a major limitation in that they can only support an exact-match search. P2P networks need to know the exact key of a data item to store that item in the responsible node. Because the exact key is given by a hash function, the key has no order relation, and the user cannot search flexibly such as when processing a range query or a multi-dimensional query.

Table 1: Node Information

Node (NodeID)	temp	Node (NodeID)	temp
N_0 (5)	21	N_4 (4)	NULL
N_1 (3)	30	N_5 (2)	29
N_2 (1)	15	N_6 (7)	18
N_3 (0)	27	N_7 (6)	24

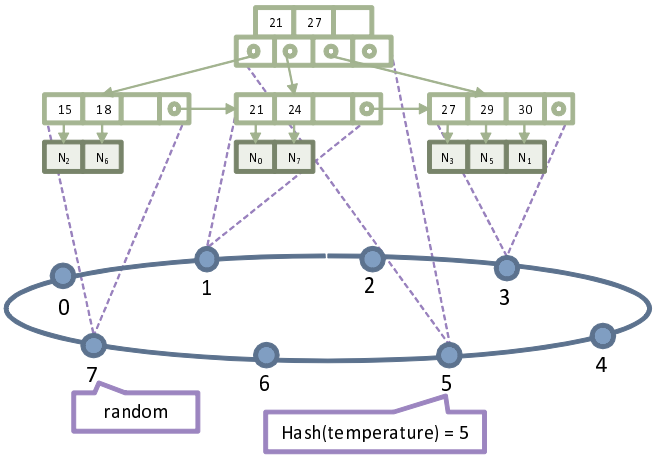


Figure 2: Mapping to P2P Networks

In P2P networks, the additional idea of searching flexibly must be applied and must often use location information. Znet [5] uses Z-ordering of space-filling curves [6] to partition the 2-dimensional ID space of the location and map space ID onto corresponding nodes, and it uses a Skipgraph [7] to manage the network topology. Mill [8] uses the same partitioning and mapping technique, but it uses Chord to manage the network topology. LL-Net [9] partitions a space into 4-blocks recursively, and the quad tree is used to manage the network topology.

These works can process a range query by using location information and also can process a multi-dimensional query by adding other properties to the location ID space as an additional dimension. However, these works have a limitation that involves including the location properties as a key in the query. In other words, these works could not execute a reverse resolution of the location information and also could not deal with the varying number of property without a reboot of the whole system because they assume that the number of dimensions, which represents properties, is static.

3 The Design of Multi-dimensional Algorithm

3.1 Overview

We are designed multi-dimensional algorithm in our previous works, and this algorithm uses a B+tree [10] for an efficient search with an arbitrary number of sensing data type. Our algorithm introduces the idea of building as many tree structures as there are properties to process a dynamic multi-dimensional range search, and these tree nodes (tree-nodes) are mapped to the node on P2P networks (P2P-nodes). The

Algorithm 1 Put(targetHash, propertyValue, timeStamp)

Require: targetHash is closest to myNodeID
  $storedValue \leftarrow SearchLocal(targetHash)$
  if $storedValue = null$ **then**
    $PutLocal(targetHash, propertyValue, timeStamp)$
  else
    if $!storedValue.isLeaf()$ **then**
      for all c such that $storedValue.getChild()$ **do**
        if $(c.min \leq propertyValue)\&\&(propertyValue < c.next.min)$ **then**
          $Put(c.getNodeID(), propertyValue, timeStamp)$
        end if
      end for
    else
      if $storedValue.isFull()$ **then**
        $Separate(storedValue)$
      end if
      $PutLocal(targetHash, propertyValue, timeStamp)$
    end if
  end if

user can select an arbitrary number of trees to process a multi-dimensional search, and a conclusive result is obtained by merging all the results. Therefore, the reverse resolution about certain properties like location can be done by using the composition query, and the system can deal with the increasing and decreasing properties while the system keeps the running state.

To build a strong tree structure for frequently varying data value such as sensor data, our algorithm uses a B+tree, which is a balanced tree, and the order t provides $O(log_t N)$ search cost, where N is the number of peers in the system. The B+tree acts as a logic structure that is used to make a comparison across property values. A network topology uses P2P to store the tree-node. Mapping from the tree-node to the P2P-node is necessary. In our algorithm, this mapping uses a hash function. Mapping is randomly done in most tree-nodes, but the root of the tree should be uniquely known to all nodes because the search for the B+tree should get the root in the beginning. Therefore, the root has a mapping rule wherein the map from the tree-node to the P2P-node of $NodeID = Hash(propertyname)$ as the manager node. On the basis of this analysis, when the nodes of Table 1 that shows a set of nodes and their property join the P2P networks, mapping is done like it is in Fig. 2. Of course, this algorithm will deal with more properties by building a new tree structure.

3.2 Algorithm

If a new node wants to join a P2P network, the node should insert its shared resource information to a tree-node on the P2P-node, such as properties and its own NodeID. This put process is done as Algorithm 1. The node that wants to insert data gets a hash value based on property name and sends an insert request to the manager node that has the same hash value. The received node relays to the child tree-node that includes the request key between the tree range. On the other hand, the node creates a root for the corresponding tree if the tree-node does not exist. By repeating this recursive opera-

Algorithm 2 Remove(targetHash, propertyValue)

Require: targetHash is closest to myNodeID
  $storedValue \leftarrow SearchLocal(targetHash)$
  if $storedValue! = null$ **then**
    if $!storedValue.isLeaf()$ **then**
      for all c such that $storedValue.getChild()$ **do**
        if $(c.min \leq propertyValue)\&\&(propertyValue < c.next.min)$ **then**
          $Remove(c.getNodeID(), propertyValue)$
        end if
      end for
    else
      if $storedValue.contain(propertyValue)$ **then**
        $RemoveLocal(targetHash, propertyValue)$
      end if
    end if
  end if

Algorithm 3 Get(targetHash, min, max)

  $treeNode \leftarrow dht.get(targetHash)$
  $result \leftarrow \{\}$
  for all e such that $treeNode.getElements()$ **do**
    if $treeNode.isLeaf()$ **then**
      if $(min \leq e.key)\&\&(e.key \leq max)$ **then**
        $result \leftarrow result \cup e$
      end if
    else
      if $(min \leq e.max)\&\&(e.min \leq max)$ **then**
        $r \leftarrow Get(e.hash, min, max)$
        $result \leftarrow result \cup r$
      end if
    end if
  end for
  return $result$

tion, the node in the target tree finally finds the destination leaf and inserts the data into the appropriate place.

The remove algorithm is shown in Algorithm 2, and large part of it is constructed in a like way of the put algorithm. There are difference in condition expression of the algorithm.

The get algorithm is shown in Algorithm 3. The algorithm also searches tree-nodes like an put and remove algorithm, but this algorithm uses a range key and an iterative search. There are two types to support the range key search in comparing the search request with its own tree-node's information. Case 1 is a comparison of the leaf; this selects the data that is included between the request range. Case 2 is a comparison of the other place; this selects the node that includes the request range between its own tree range as the next candidate. The algorithm repeats case 2 to find some destination leaf-nodes, and it executes the case 1 to get the result. We used an iterative search to repeat case 2 because a recursive search risks a fatal performance decrement in the environment of our target, such as the general middle-ware for a WSN management system. The tree-node needs to split the request in the same number of branches if two or more candidates are found, and then the tree-node must wait for the under-layer processing to finish or timeout in accordance with a recursive search. The iterative search can prevent the risk because it forces waiting time on the node that sent the search request.

3.3 The Problem of Previous Work

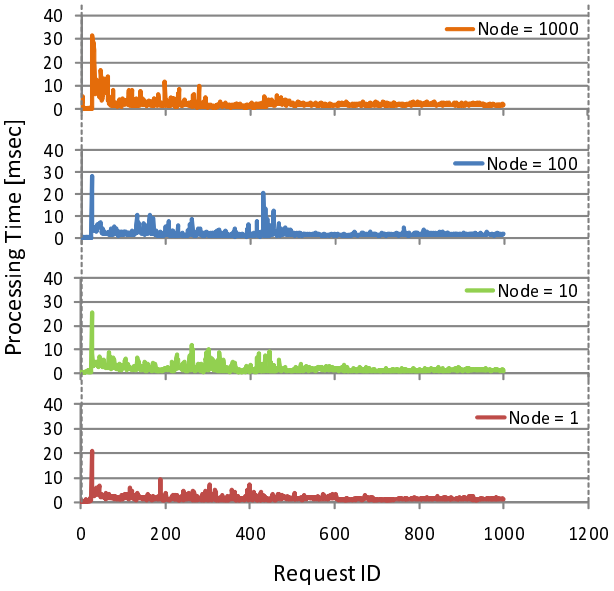
The actual sensor data occurs with great frequency because the sensor is generally used to get the realtime and high-density data. The first thing to do in the algorithms is to accsess the root of tree structure and the root intends to have a heavy workload than other nodes. Therefore, the root of tree structure may become the bottleneck in the system. A bottleneck arises from the insufficient disk space and memory, the not enough capacity of CPU, and the exclusive control method.

- The insufficient disk space and memory
  When data number that push the envelope of present node is inserted, the node frequently causes the swap and the performance decreases. In addition, the node that too many data is concentrated caused a serious performance decrease in search stored data. They become the problem in the general relational database(RDB), but Key-Value store(KVS) on P2P is not really acknowledged as a problem. The KVS can easily deal with the increasing data by distributing data to many node, and this is called scale-out technology. Therefore, this factor is not considered in this paper.
- The not enough capacity of CPU
  When the node receives the requests, CPU utilization and processing time is increases according to the request type and the current state, and it becomes the problem when the number of request increases up to a certain number. The typical KVS can deal with this problem because they can distribute not only data but also load of the processing request. However, the multi-dimensional algorithm on KVS cannot deal with the problem well because it constructs the large tree structure. This structure can support the tabular form to KVS and is guranteed to reach the target, but it limits the route to a certain data and causes a local load concentration. In particular, the node that included in the route, like the root, have a workload than other.
- The exclusive control method
  This factor is the endemic problem of distributed system. The KVS used by quite a lot of users, and the operation demand may occur around the same time through the request. The structure breaks if operation is executed in parallel because it constructs the tree structure as previously mentioned. To address the problem, the exclusive control method such as lock is needed. However, this method has a bad influence at the processing time of the request per second. In other words, it is important to reduce the maximum number of concurrent connection to keep the throughput.

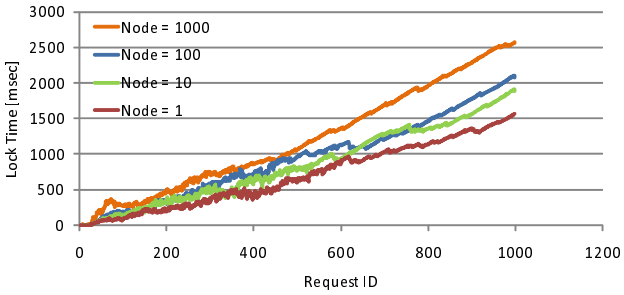
Based on above analysis, it is necessary to evaluate the time of the process and the lock before it applies to the WSN management system. We think the through-put of the system is necessary as many as traditional RDB to adopt as the management system.

4 Performance Evaluation

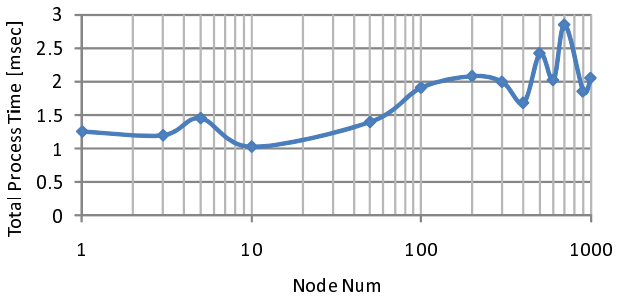
In this section, we describe evaluation of our algorithm in some delay time. The time has 3 types, and it is the com-



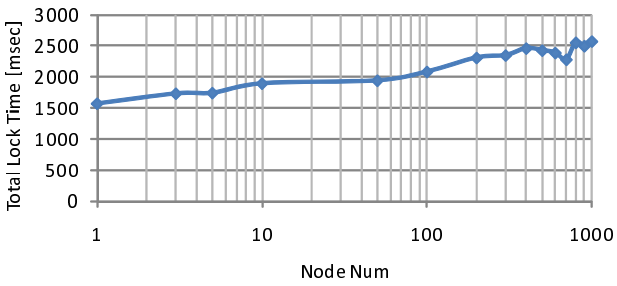
(a) Processing Time



(b) Lock Time



(c) Total Processing Time



(d) Total Lock Time

Figure 3: The Result Varying Number of Node

Table 2: Simulation Environment

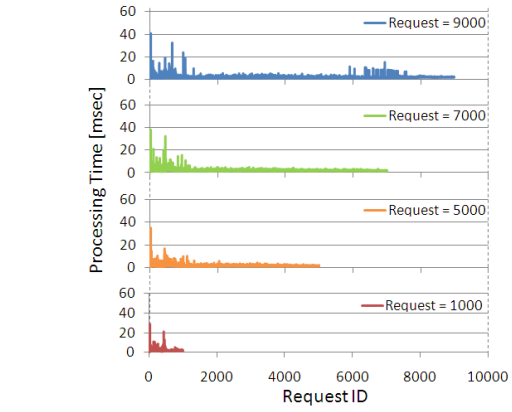
Varying	Node	Data
Num Nodes	1 – 1000	100
Num Request	1000	1 – 10000
Data Distribution	Normal	Normal

munication time, the lock time, and the processing time. The communication time evaluate the time during sending and receiving of the request, but this time is not evaluate because of using simulation. The lock time evaluate the time during receiving and processing start. The processing time evaluate the time during processing of request. We evaluate the lock and the processing time when some request is sent to one root node just around the same time. The evaluation was conducted via simulations using Overlay Weaver [11] in the environment shown in Table 2, the list size of the tree-node in our algorithm was adjusted in the experiment to 25, and algorithm use Chord to construct the base P2P networks.

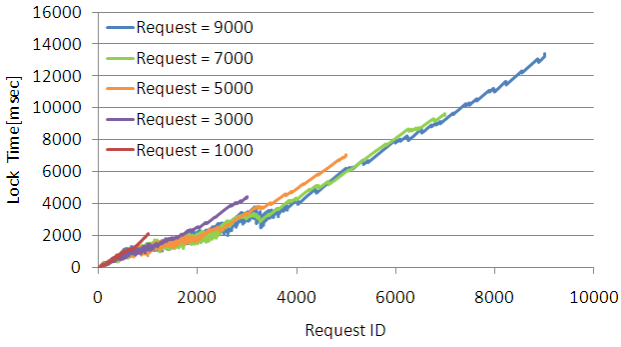
Fig. 3(a), 3(b) show the history of processing each request that have a sequential unique ID, where Node is the number of P2P network. Fig. 3(c), 3(d) show the total time for processing the last one at each number of node. Fig. 4(a), 4(b) also show the history of processing each request that have a sequential unique ID, where request is number of inserted request at the same time. Fig. 4(c), 4(d) show the total time for processing the last one at each number of node.

Fig. 3(a) has some part where processing time increases suddenly. This reason is that the B+tree structure must divide a leaf that is full, create new leaf to store the leaf information, and may increase the number of layer on the tree structure in addition to a usual operation. These additional operation occur when the number of leaf becomes more than certain number that is 25^i in an ideal environment, where 25 is adjusted as the leaf size. The additional operation also affects to Fig. 3(b). The part near the origin point in Fig. 3(b) is very low delay time, and the time suddenly grows when 25 request ID is exceeded because the root can process the request only for myself until the 25. When the Fig. 3(c), 3(d) is paid attention, the influence of the node number is very little because the order is $O(logN)$, where N is the number of nodes. The reason of this order is that we constructed the tree structure on P2P of Chord algorithm. Chord algorithm search cost is $O(logN)$, we use Chord search method to find the appropriate child node on the tree structure, and then the influence of Chord search order appeared to the result. We are sure that our idea, building the large structure on P2P to handle a multi-dimensional as our work, is the enough scale-out.

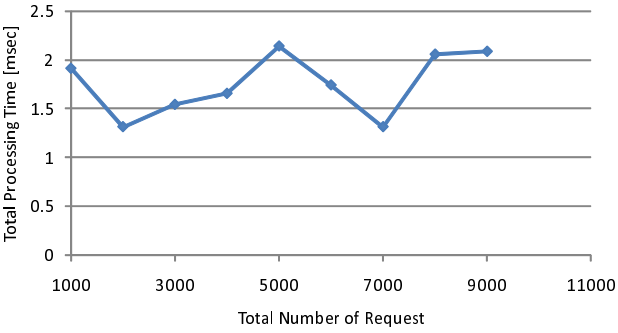
Fig. 4(a) have similar change with Fig. 3(a), and the change of Fig. 4(c) is smooth because the rapid change occurs by dividing and creating the leaf when the leaf is full. In Fig. 4(b), the varying total number of request is not affected the lock time, and the result is linearly increases as shown in Fig. 4(d) because the request go into a queue once even if the system receive any amount of request and any time. On the other hand, Fig. 4(d) shows through-put that can process the x request by y msec from a point of view of the system. We can obtain the through-put that absolutely guaranteed to process



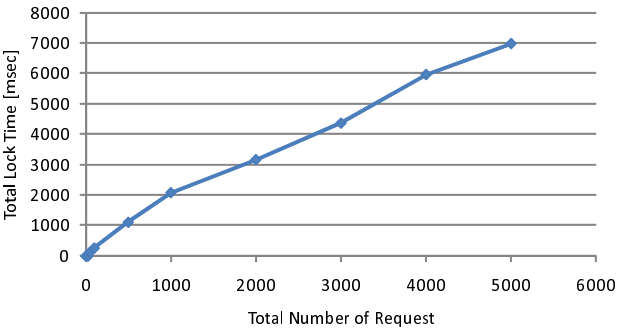
(a) Processing Time



(b) Lock Time



(c) Total Processing Time



(d) Total Lock Time

Figure 4: The Result Varying Number of Request

as soon as receiving from the intersection in figure with expression $y = 1000$, and it is about 500 requests per second. This through-put is not too few because the traditional RDB's default number of concurrent connection are from 100 to 500 on demand and the some system often remain of the default number.

5 Conclusion

We designed a the multi-dimensional search algorithm in P2P networks using a B+tree. Our novel P2P index structure is well suited for applications such as a sensing data sharing system by supporting range and multi-dimensional queries. This structure is in accordance with the basic key idea that a tree structure builds on P2P networks for each property such as a temperature-tree. We evaluated the performance of the proposed method by simulation. From simulation results, we found that the proposed algorithm has scalability because the processing and lock time is order $O(logN)$. In addition, the through-put that guaranteed to process the request soon is about 500 requests per second. In the future, we will aim at the performance gain of the system and construct a system using real sensor data on real networks.

REFERENCES

[1] M. Ceriotti, L. Mottola, G. Picco, A. Murphy, S. Guna, M. Corra, D. Zonta, and P. Zanon, "Monitoring Heritage Buildings with Wireless Sensor Networks: The Torre Aquila Deployment", Proc. IPSN, pp. 277 – 288, 2009.

[2] J. Kim, J. Lim, J. Friedman, U. Lee, L. Vieira, D. Rosso, M. Gerla, and M. Srivastava, "SewerSnort: A Drifting Sensor for In-situ Sewer Gas Monitoring", Proc. IEEE SECON, pp. 691 – 699, 2009.

[3] G. Gutierrez, B. Mejias, P. Roy, D. Velasco, and J. Torres, "WSN and P2P: A Self-Managing Marriage", Proc. IEEE SASO, pp. 198–201, 2008.

[4] N. Matsuura, H. Mineno, N. Ishikawa, and T. Mizuno, "Evaluation of B+Tree-based Multi-dimensional Range Search Algorithm for P2P Networks", Proc. IEEE SAINT, 2010.

[5] Y. Shu et al. "Supporting Multi-dimensional Range Queries in Peer-to-Peer System", Proc. IEEE P2P, pp. 173 – 180, 2005.

[6] J. Wierum, "Logarithmic Path-Length in Space-Filling Curves", Proc. CCCG pp. 22 – 26, 2002.

[7] J. Aspnes et al. "Skip Graphs", Trans. ACM Algorithms, vol. 3, no. 4, p. 37, 2007.

[8] S. Matsuura et al. "Mill: Scalable Area Management for P2P Network based on Geographical Location", Proc. Euromedia, pp. 46 – 52, 2006.

[9] Y. Kaneko et al. "A location-based peer-to-peer network for context-aware services in a ubiquitous environment", Proc. SAINT, pp. 208 – 211, 2005.

[10] D. Comer, "The Ubiquitous B-Tree", ACM Computing Surveys, vol. 11, no. 2, pp. 121 – 138, 1979.

[11] Overlay Weaver, http://overlayweaver.sourceforge.net/.

Evaluation of a novel space filling curve and its application to the P2P overlay network

Tomoya Kitani†, and Yoshitaka Nakamura‡

†Division of Global Research Leaders, Shizuoka University, Japan
‡Graduate School of Information Science and Technology, Osaka University, Japan
t-kitani@ieee.org, y-nakamr@ist.osaka-u.ac.jp

Abstract - P2P networks can achieve high scalability since they distribute service contents/resources to multiple nodes in the network. In a P2P network, it is necessary to search the resource location on the network when we use some contents/resources. Space filling curve is known as technique to map information of a multi-dimensional space such as location information onto a one-dimensional space such as ID. We have proposed a novel space filling curve for P2P overlay networks considering the proximity of nodes based on the geographical information of the nodes and reducing the communication delay proportional to the geographical distance of the nodes. By using the proposed space filling curve, we can convert geographic information of nodes into their ID (label) of P2P network. In this paper, we evaluate the space filling curve for P2P networks and apply it on a virtual network map. On the virtual network map, each node has the coordinated position information which is proportional to the communication delay with its neighbor nodes. Through the numerical evaluation, we confirmed that the proposed curve is more suitable for handling hierarchical-spread nodes than the conventional curves on the virtual network map.

Keywords: P2P, Overlay Network, Space Filling Curve, Virtual Network Map, Node Labeling

1 Introduction

In recent years, with the development of the communication technology, mobile terminals such as cellular telephones or the note PCs come to be connected frequently to the network with wireless communication. As a result, the user of mobile terminals become possible to receive e-mails or the multimedia contents such as the image, and movie, etc. without depending at the specific location and time. Moreover, a huge amount of terminals such as sensors deployed in the living space or terminals installed by cars are expected to compose the network in coming ubiquitous computing society.

It is difficult to keep a scalability to offer high quality service for such large amount of terminals because the load to the server concentrates in server-client type system mainly used in current computer network. Therefore, the Peer-to-Peer (P2P) network system attracts attention. In the P2P network system, each terminal manages data autonomously and distributedly, and communicates directly with other terminals. Gnutella[2] and BitTorrent[1] are typical P2P network technologies. P2P network can provide network services to millions of terminals by distributing the load of the terminals and network unlike the server-client type architecture.

On the other hand, in the ubiquitous computing society, it becomes available a location-aware service that selects appropriate services by using user's geographic information, and a contents-aware service that selects appropriate services considering of the state of sensors in the neighborhood. Current cellular phone terminals and many of in-vehicle navigation systems are equipped with GPS (Global Positioning System) and can easily acquire the geographical location information of each user.

Space filling curve is known as technique to map information of a multi-dimensional space such as location information onto the one-dimensional space such as ID. Location-aware services can be provided easily by using location information as a node ID on P2P networks. However, there is a trade-off between the inefficiency of space filling in simple method such as Z-Ordering (Lebesgue curve)[10] and high calculation cost of a curve with good filling efficiency.

We have proposed a novel space filling curve for P2P overlay networks considering the proximity of nodes based on the geographical information of the nodes and reducing the communication delay proportional to the geographical distance of the nodes [8]. The proposed method gives each node a label (ID) from the space coordinate of the node based on its geographic information and the link delay between the nodes. The proposed curve can let the space coordinate be converted into the one-dimensional space efficiently. In this paper, we evaluate the space filling curve for P2P networks and apply it on a virtual network map by Vivaldi[4]. On the virtual network map, each node has the coordinated position information which is proportional to the communication delay with its neighbor nodes.

The rest of this paper is organized as follows. In Section 2, we briefly introduce related work of construction methods of P2P networks. Next, we describe our space filling curve and its application to a virtual network map in Section 3. Then we present neumerical simulation results in Section 4. Finally, we conclude this paper in Section 5.

2 Related Work

In Peer-to-Peer(P2P) technology, each peer constructs application layer overlay network, searches contents or resources over this network by direct communication between peers without a server, and distributes and shares the resources. Gnutella[2] and BitTorrent[1] are typical technologies using the non-structured overlay networks. On the other hand, the structured overlay networks can search a resource on the P2P networks, and many technologies are proposed in recent years.

Especially, there are many technologies that use Distributed Hash Table(DHT). Chord[13], CAN (content addressable network)[11], Pastry[12], and Tapestry[15] are known well.

On the other hand, in order to enable the range search which treats the consecutive quantity, some techniques which do not use DHT for decentralization of data and search queries have been proposed. SkipNet[5] is one method for configuring a structured P2P overlay network. In this method, the range search is enabled by using SkipGraph[3] that uses consecutive values for ID instead of DHT that uses a hash function. SkipGraph achieves the search efficiency of the logarithmic order by hierarchically grouping one-dimensional node array using the balanced tree structure. However, the range specification search in SkipNet might need to search wider range than other methods, because SkipNet constructs a hierarchical overlay network by using single search key, without regard to the geographical location and multiple search keys. LL-net[6] is structured P2P overlay network where the area on the map is hierarchically divided into four sub areas, and in each hierarchy the overlay links should be the different length links. In LL-net, the scalability of system is lacked so that the existence of a special node that manages each area may be assumed, and the management cost of the overlay network may increase. Moreover, if the distribution of the node is not uniform, it might become impossible to converse from geographic coordinates into node ID easily, the hierarchical structure might become biased, and the search efficiency might turn worse. In Mill[9], the range search is enabled by connecting nodes with the ring structure by applying ID which based on geographical location information. Hereby, the search efficiency can be realized $O(logN)$ without special nodes.

Space filling curve[14] (Peano curve) is known as a technique to convert information on multi-dimensional space such as location information into information on one-dimensional space like ID. The typical space filling curves are Lebesgue curve (Z-Ordering, Fig.1), Hilbert curve (Fig.2), Sierpinski curve (H-indexing), and $\beta\Omega$-indexing, etc.

Vivaldi[4] is a technique to calculate the coordinate system autonomously by considering both the proximity of the physical network and real communication delay on a P2P network. This technique leads virtual coordinate system by gradually correcting the difference between Euclidean distance and the actual measurement delay between nodes using the spring model after each participation node decides its virtual coordinates autonomously.

3 Proposed Method

In this paper, we research a new configuration method for structured P2P overlay network considering delay variations between nodes. To match coordinates in the two-dimensional plane with delay variations of nodes we use the technique of Vivaldi when the coordinates are calculated.

The search efficiency can be improved when given an ID which has small distance from IDs of the physically neighboring nodes. Therefore, it can be good for labeling a node ID to use geographical information of nodes. The space filling curve is often used for this case. We have proposed a space filling curve in [8] with following features that (1) it is easily convertible on node ID on the P2P network from the geographic information, (2) information search that specifies the range is efficiently executable, and (3) it is easy to construct a hierarchical structure to give the scalability.

In this paper, we describe the method to let the proposed curve fit the delay variations of nodes.

3.1 Conventional Space Filling Curves

Some space filling curves have been proposed so far. Hilbert curve, Sierpinski curve, and $\beta\Omega$-indexing fill the space so that the neighbor nodes on the two-dimensional space is converted to the close ID on the one-dimensional space. In these curves, the communication delay between logical neighbor nodes is suppressed small, and the range search can be searched efficiently. The performance of the Lebesgue curve is not better than that of the above curves, because the terminal nodes of the link that connects between clusters are a long way each other on the two-dimensional space (see Fig. 1). However, Lebesgue has simple structure and two dimensional coordinates can be easily converted into the corresponding IDs on it. Therefore, Lebesgue curve is often used for ID conversion in P2P networks, especially for mobile ones where the computing power is limited and it is required to reduce overhead of complex processing. Our space filling curve aims to be able to convert ID as easily as the Lebesgue curve and to give an ID the correlation with the geographical proximity of the node which has the next ID such as Hilbert curve.

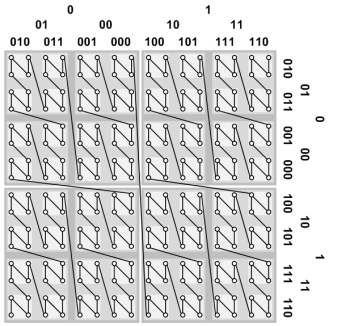


Figure 1: Lebesgue Curve (Z-Ordering)

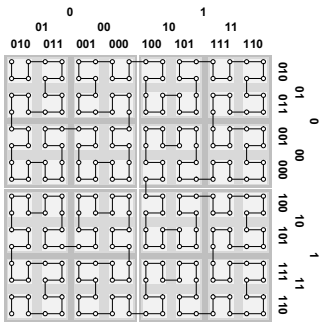


Figure 2: Hilbert Curve

3.2 Consideration on the Curves and Proposed Curve

The conventional space filling curves such as Hilbert curve and $\beta\Omega$-indexing have self-similarity because they are constructed by hierarchical processes. But they are not considered to design to hierarchically spread nodes. In the curves, it can be possible to change the hierarchical level with respect to each sub area, for example there are 2^4 nodes to be traveled (the hierarchical level is 2) in an sub area and there are 2^6 nodes to be traveled (the hierarchical level is 3) in another sub area. However, in the sub areas, all nodes should be in the same hierarchical level and they should be relabeled when the number of nodes increased. It is difficult for the

labeling scheme of the curves to be assigned hierarchically-spread nodes efficiently. In order to achieve scalability for information storage and retrieval on P2P networks based on the curves, service providers need to adopt a hierarchical structure in an upper service layer. The curve to be proposed can be flexibly assigned and label hierarchically-spread nodes in each area. It can travel all nodes on the curve easily. The curve can handle hierarchical-spread nodes and achieve scalability at the labeling scheme.

Only Lebesgue curve could be assigned such nodes by enhancing the curve to assign hierarchical nodes at the middle point of each edge which goes across an area vertically. In the evaluation section, we compare our curve with the enhanced Lebesgue curve.

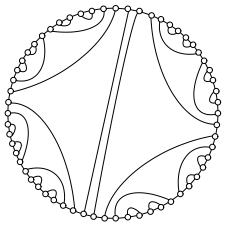


Figure 3: Hierarchical Chordal Ring Network ($N = 62$)

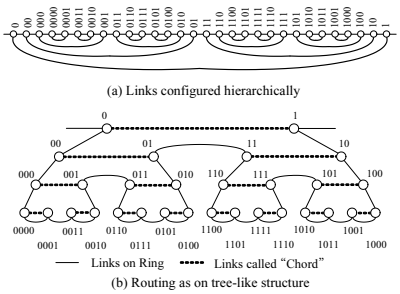


Figure 4: A cluster of HCRN

3.3 Constitution of the Space Filling Curve

In this section, we explain about the proposed space filling curve[8]. The proposed space filling curve is composed based on Hierarchical Chordal Ring Network (HCRN)[7] shown in Fig.3 by enhancing HCRN to two dimensions.

HCRN constructs the tree structure topology on the ring by connecting the link on the node on a cluster edge after clustering by recursive division of node clusters on the ring (Fig.4). Here, each node's ID is expressed by a variable-length gray code according to the hierarchy where the node is located, and each node queues up in the order of the gray code in the hierarchy. The binary number and the gray code can be easily converted by using the following exclusive-OR operation and the simple algorithm. If a certain n digit binary number is assumed to be $\mathbf{b}$ and the gray code corresponding to it is assumed to be $\mathbf{g}$, they are possible to be converted to each other.

Algorithm Binary2Gray

$$\mathbf{g} = \mathbf{b} \oplus (\mathbf{b} >> 1)$$

Algorithm Gray2Binary

1: Given a gray code $\mathbf{g} = (g_1, g_2, \cdots, g_n)$
2: $flag_{rev} \leftarrow$ false
3: **for** $i \leftarrow 1$ **to** n **step** 1
4: **if** $g_i =$ '1' **then**
5: **if** $flag_{rev} =$ false
6: **then** $b_i \leftarrow$ '1' **else** $b_i \leftarrow$ '0' **endif**
7: $flag_{rev} \leftarrow$ **not** $flag_{rev}$
8: **else**
9: **if** $rev_flag =$ false
10: **then** $b_i \leftarrow$ '0' **else** $b_i \leftarrow$ '1' **endif**
11: **endfor**
12: **return** the binary code $\mathbf{b} = (b_1, b_2, \cdots, b_n)$

In Lebesgue curve, given a geographical position $(\mathbf{x_i}, \mathbf{y_i})$ of a node n_i $(\mathbf{x_i} = (x_1, x_2, \cdots, x_H), \mathbf{y_i} = (y_1, y_2, \cdots, y_H))$, the node's geographical label $\mathbf{p_i}$ is given as $\mathbf{p_i} = (x_1, y_1, x_2, y_2, \cdots, x_H, y_H)$. Here, H denotes both the length of geographical information and the maximum number of hierarchical level.

In our curve, first, in order to let the curve be a closed curve, we define that a new geographical information of a node n_i is given as $\mathbf{p_i} =$ Binary2Gray$(x_1, y_1, \overline{x_2}, \overline{y_2}, \cdots, x_{2j-1}, y_{2j-1}, \overline{x_{2j}}, \overline{y_{2j}}, \cdots)$. Here, $\overline{x}$ means bit inversion and $j \in \mathbb{N}$. This definition provides a result that the nearer two nodes is located, the smaller the Hamming distance between their node labels is.

Next, to handle hierarchically-spread nodes, we define the following algorithm to label ID to each node. Each node is labeled by this scheme in the chronological order of the nodes.

Given

$\mathbf{p_i} = f(\mathbf{x_i}, \mathbf{y_i})$: The geographical label of each node n_i
H : Maximum hierarchical level

Procedure

1: Given $\mathbf{p_i} = (p_1, p_2, p_3, \cdots, p_{2H})$
2: **for** $j \leftarrow 2$ **to** $2H$ **step** 2
3: **if** Label $(p_1, \cdots, p_j)$ has not been used yet
4: **then** Label $\mathbf{l_i} \leftarrow (p_1, \cdots, p_j)$ and **goto** 7.
5: **endif**
6: **endfor**
7: end.

The enhanced Lebesgue curve is obtained by connecting nodes in the numerical order of the label $\mathbf{l_i}$. Note, however, suppose $(l_1, l_2, \cdots, l_m, 0, 1) < (l_1, l_2, \cdots, l_m) < (l_1, l_2, \cdots, l_m, 1, 0)$ when the length of the labels are different.

The proposed curve is obtained by connecting nodes in the numerical order of Gray2Binary$(\mathbf{l_i})$. Note, however, suppose $(l_1, \cdots, l_{m-1}, 0) < (l_1, \cdot, l_{m-1}, 0, l_{m+1}, \cdots, l_n)$ and $(l_1, \cdots, l_{m-1}, 1, l_{m+1}, \cdots, l_n) < (l_1, l_2, \cdots, l_{m-1}, 1)$ when the length of the labels are different.

An example of the proposed curve is shown in Fig. 5. Figure 6 shows the lowest level nodes and their connections out of the curve in order to let it be easier to compare with the conventional space filling curves such as Hilbert curve.

3.4 Applying to Virtual Network Map

3.4.1 Overview

We use the proposed space filling curve to calculate IDs of nodes from the geographical coordinate of each node uniformly and configure P2P network. But, if nodes are unevenly distributed, some problems occur such that the load concentrates on specific nodes, and that the balance of network topology is collapsed and the diameter of network becomes large. Because there is positive correlation in the communication delay in the network and the geographical proximity between arbitrary nodes, LL-net makes nodes a cluster

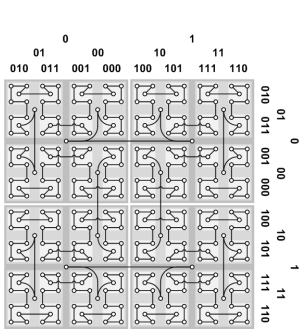


Figure 5: Proposed curve

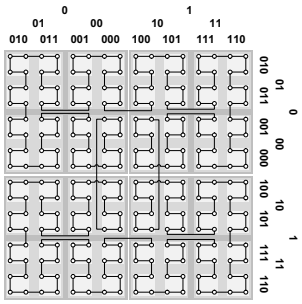


Figure 6: The links of the lowest level in the proposed curve

using geographical information to construct low-delay P2P network. Moreover, it is possible to deal with the problem of the bias of the load of each node and the increase of network diameter, because the multi-level network is dynamically made for geographical uneven distribution of nodes. In the large network that the distances between nodes are hundreds of kilometers-scale, the geographical proximity is thought to have great influence on the communication delay. However, the influence does not grow so much in the network that the distances between nodes are kilometers scale. Rather, the increase of the adjustment cost for the disorder of the hierarchical structure by the uneven distribution of the node becomes the big problem. In the foregoing paragraph, it have been described that the proposed space filling curve can consider the geographical proximity and the proximity in the network between nodes more by almost the same ID conversion cost as the Lebesgue curve. In the network with many mobile terminals, to reduce the ID conversion cost becomes more important than to minimize the real communication delay of the connection link, from the viewpoint of low ability for computing / communication and mobility of the mobile terminals. Therefore we propose the configuration method of the P2P overlay network which lower the communication delay in the network and reduce the problem of uneven distribution of nodes under the supposition that each node can acquire by information about the communication delay.

3.4.2 Construction of Virtual Network Map

At first, in the proposed method, each node to build an overlay network is given a virtual coordinate with a virtual network map. We use Vivaldi[4] as configuration technique of the virtual network map in this method. Hereby, the virtual coordinates are set by considering real delay between nodes and nodes are not concentrated at the specific location on the virtual network map. Next, node ID is assigned to each node in the virtual network map by applying the proposed space filling curve.

In this method, we use Vivaldi to give the coordinate of each node on the virtual network map. Vivaldi is technique to build a coordinate system for autonomous decentralized in each node by coordinating the error of actual delay value with Euclid distance using the principle of the spring gradually. However, there are problems that the convergence of the coordinate system is slow because width of the adjustment of the error is small. Therefore, the prolongment of the settling time becomes the problem especially when the coordinates of the initial solution are quite different from an actual delay value. In our method, the geographical location information are given as initial coordinates. We assume that each node which participate in the network can acquire location information by GPS. Each node uniformly calculates initial coordinate from the acquired location information. Figure 7 shows the outline of Vivaldi.

The P2P overlay network that considers the delay time and the uneven distribution of nodes can be constructed by calculating node ID with the proposed space filling curve for the nodes on the virtual network map with Vivaldi.

4 Numerical Evaluation

The advantage of the proposed method is evaluated by the following points.

- The index that shows the physical neighbor node is also near on space filling curve (Figs. 9 and 10)
- The communication delay which necessary for travelling all over the filling curve (Figs. 11 and 12)

First, we compared the proposed curve with Lebesgue curve and Hilbert curve form the standpoint of the proximity index. Next, we compared it with the enhanced Lebesgue curve. Nodes on the proposed curve and the enhanced Lebesgue curve can belong to various hierarchical levels as shown in Fig. 5 although nodes on Hilbert curve belong to the same hierarchical level.

4.1 Proximity with Neighbor Nodes on the Curves

First of all, we evaluated how far the nodes which is physically in four neighborhoods which are on the filling curves, using logarithmic index range R_{avg}^{curve} shown in [14]. Suppose that each node n_i has a sequence number $seq(i)$ on the curve and it is located on coordinate $pos(i) = (x_i, y_i)$. Proximity index R_{avg}^{curve} can be calculated by the following expressions.

$$
\begin{aligned}
r_1^{curve}(i) &= \log(|seq(i) - seq(pos^{-1}(x-1, y))| + 1) \\
r_2^{curve}(i) &= \log(|seq(i) - seq(pos^{-1}(x+1, y))| + 1) \\
r_3^{curve}(i) &= \log(|seq(i) - seq(pos^{-1}(x, y-1))| + 1) \\
r_4^{curve}(i) &= \log(|seq(i) - seq(pos^{-1}(x, y+1))| + 1) \\
R_{avg}^{curve} &= \mathrm{avg}_i\{r_j^{curve}(i)\}, \quad (j \in \{1, 2, 3, 4\})
\end{aligned}
$$

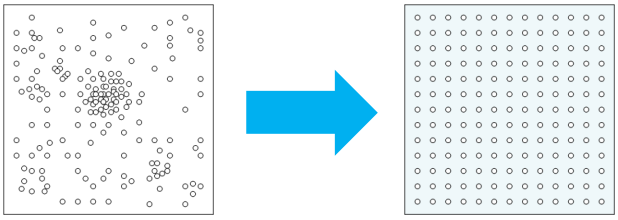


Figure 8: Construction of Virtual Network Map by Vivaldi

- Input
  - N : Number of nodes
  - $Nodes = \{n_i\}(i = \{1, 2, \cdots, N\})$: The set of the nodes
  - $\mathbf{x_i} = (x_i, y_i)$: The geographical position attribute of each node n_i
  - $D = \{d_{ij}\}$: The matrix of delay time d_{ij} between arbitrary nodes n_i and n_j
- Output
  - $\mathbf{p_i} = (p_i, q_i)$: The position of each node n_i on the delay map
- Objective function $cost$
  - To minimize E which is the total sum of the estimation error between communication delay and distance on the delay map

$$\text{minimize } \Sigma_{i,j}(d_{ij} - ||\mathbf{p_i} - \mathbf{p_j}||)^2$$

(a) Formulation

- F : Force vector that the spring between node i and j exerts on node i
- $u(i, j)$: Unit vector which gives the direction of the force between i–j
- t : The time constant to converge the spring model

While $(cost - cost_{prev}) < \epsilon$
  foreach $i \in Node$
    $F := 0$;
    foreach $j \in Node$
      $e := d_{ij} - ||n_i - n_j||$;
      $F := F + e \times u(i, j)$;
    end;
    $\mathbf{p_i} := \mathbf{p_i} + t \times F$
  end;
end.

(b) Algorithm

Figure 7: Overview of Vivaldi

The evaluation results are shown in Fig.9.

A physically neighbour node can be assigned to nearer sequence number on the proposed curve comparing with the Lebesgue curve. It can be a remarkable improvement, especially considering that R_{avg}^{curve} is the log scale. Although Hilbert curve shows much better result, it is difficult to use Hilbert curve on the P2P network because there is no easy way for nodes on Hilbert curve to decide where the next node is.

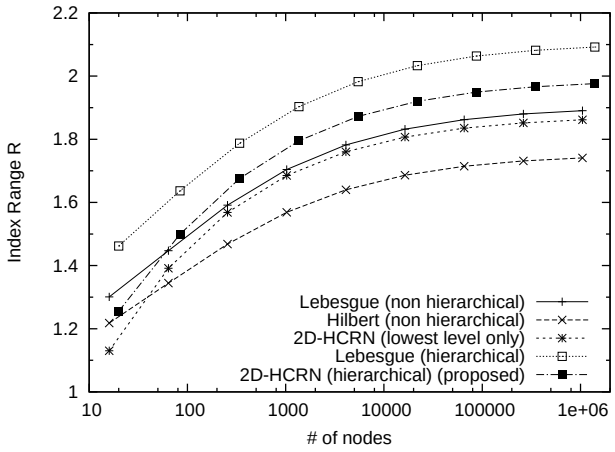


Figure 9: Average of logarithmic index range R_{avg}^{curve}

4.2 Length of Curves

We evaluated the delay for various numbers of nodes with computer simulations. In the simulations, we assume that the number of nodes are changed from 50 to 10,000, and each result is the average of 30 trials.

Figure 10 shows the square sum of delay between each node and each next node on the curves where the geographical position of nodes are generated by the uniform distribution. As a result, the proposed curve has been improved by

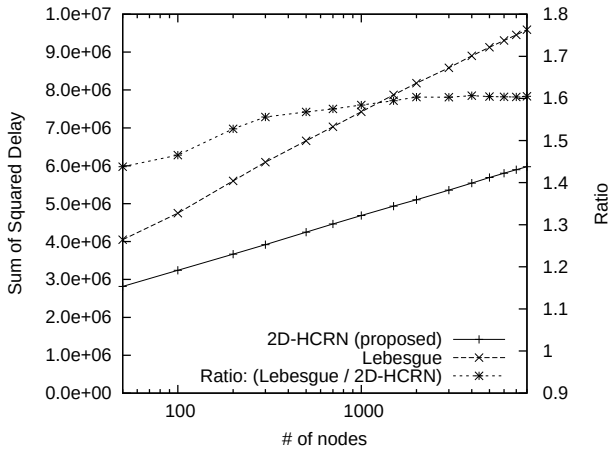


Figure 10: Sum of squared delays between successive two nodes on the curves with randomly distributed nodes

as much as 30-60% against Lebesgue curve. It can be said that the distance between any adjacent nodes on the proposed curve are always small without large bias than the enhanced Lebesgue curve.

4.3 Traversal Time on Virtual Network Map

Finally, we evaluated the traversal time of all nodes on the proposed curve and on Lebesgue curve (Z-ordering) of each when the virtual network map is used and when it is not used. The number of nodes is given between 10 to 100 in this experiment. Figures 11 and 12 show the result of simulations.

On the condition that those curve are not on a virtual network map, there is few difference between the proposed curve and Lebesgue curve because most of the delay is dominated by physical link connectivity of each node rather than the proximity of neighbor nodes on the curve. However, when the virtual network map is used, the traversal time of the proposed curve is smaller than that of Lebesgue curve because the delay between nodes is proportional to the distance between the nodes on the virtual network map. In this way, our

proposed curve is effective in the network with delay variations.

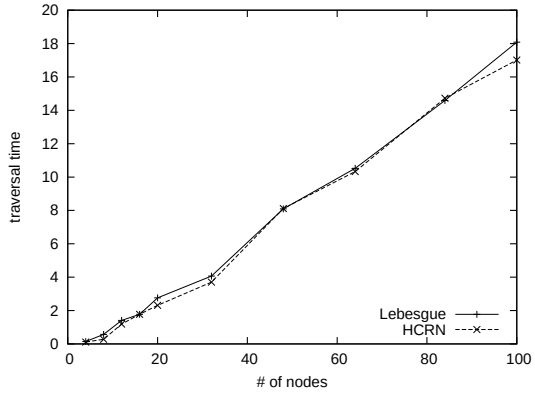


Figure 11: Traversal time on the curves not o virtual network map

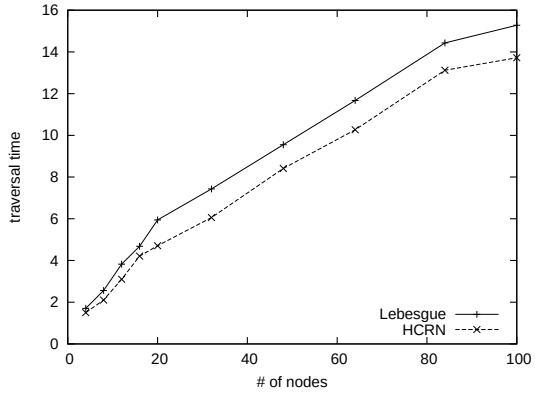


Figure 12: Traversal time on the curves on virtual network map

5 Conclusion

In this paper, we proposed the method which configure small delay structured P2P overlay network by assigning a label considering the geographic information of each terminal node that became peer. The proposed space filling curve to configure the P2P network converts geographic information into node ID (label) of P2P network, and it can search information over specified range efficiently, and can construct hierarchical connection structure for scalability. Through the numerical evaluation, we confirmed that the proposed curve is more suitable for hierarchical-spread nodes than the conventional curves.

For future work, we need the performance evaluation of the proposed space filling curve and the P2P overlay network in the real network environment especially in node distribution with bias, and the reexamination of the routing entry of each node to improve the performance of this method.

REFERENCES

[1] Bittorrent. `http://bittorrent.com`.

[2] Gnutella. `http://gnutella.wego.com`.

[3] J. Aspnes and G. Shah. Skip graphs. In *Proceedings of the 14th annual ACM-SIAM Symposium on Discrete Algorithms (SODA 03)*, pages 384–393, January 2003.

[4] F. Dabek, R. Cox, and R. Morris. Vivaldi: A decentralized network coordinate system. In *Proceedings of the ACM SIGCOMM 2004*, pages 15–26, August 2004.

[5] N. J. A. Harvey, M. B. Jones, S. Saroiu, M. Theimer, and A. Wolman. Skipnet: A scalable overlay network with practical locality properties. In *Proceedings of the 4th USENIX Symposium on Internet Technologies and Systems (USITS '03)*, pages 113–126, March 2003.

[6] Y. Kaneko, K. Harumoto, S. Fukumura, S. Shimojo, and S. Nishio. A location-based peer-to-peer network for context-aware services in a ubiquitous environment. In *Proceedings of Interjatinoal Symposium on Applications and the Internet (SAINT 2005)*, pages 208–211, February 2005.

[7] T. Kitani, N. Funabiki, H. Yamaguchi, and T. Higashino. Hierarchical logical topology in wdm ring networks with limited adm. In *Proceedings of the IFIP Networking 2008 (Networking 2008)*, pages 326–337. Springer-Verlag LNCS 4982, May 2008.

[8] T. Kitani and Y. Nakamura. A configuration method for structured p2p overlay network considering delay variations. In *Proceedings of the International Workshop on Information Technology for Innovative Services (ITIS 2009)*, pages 547–552, August 2009.

[9] S. Matsuura, K. Fujikawa, and H. Sunahara. Mill: A geographical location oriented overlay network managing data of ubiquitous sensors. *IEICE TRANSACTIONS on Communications*, E90-B(10):2720–2728, October 2007.

[10] J. Orenstein and T. Merrett. A class of data structures for associative searching. In *Proceedings of the 3rd ACM SIGACT-SIGMOD Symposium on Principles of Database Systems*, pages 181–190, April 1984.

[11] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Schenker. A scalable content-addressable network. In *Proceedings of the ACM SIGCOMM 2001*, pages 161–172, August 2001.

[12] A. I. T. Rowstron and P. Druschel. Pastry: Scalable, decentralized object location, and routing for large-scale peer-to-peer systems. In *Proceedings of the IFIP/ACM International Conference on Distributed Systems Platforms Heidelberg (Middleware 2001)*, pages 329–350. Springer-Verlag LNCS 2218, November 2001.

[13] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and Hari Balakrishnan. Chord: A scalable peer-to-peer lookup service for internet applications. In *Proceedings of the ACM SIGCOMM 2001*, pages 149–160, August 2001.

[14] J.-M. Wierum. Logarithmic path-length in space-filling curves. In *Proceedings of the 14th Canadian Conference on Computational Geometry*, pages 22–26, August 2002.

[15] B. Y. Zhao, J. Kubiatowicz, and A. D. Joseph. Tapestry: An infrastructure for fault-tolerant wide-area location and routing. Technical report, Computer Science Division, University of California at Berkeley, 2001.

Panel Discussion 1
(Chair Tomoya Kitani)

Being sustainable through Knowledge Resource Management

Kazuya Matsudaira *

* **System Frontier Corporation**
mtdr@pji.co.jp

Abstract –Any organization is said as a bio-Corporation who has been adaptable to sudden changes of economic and political environments. It might be very difficult to live in 100 years if human being. Average age of Japanese company is about 41 years old. Current number of company who live more than 100 years old is not less than 20 thousand companies in Japan [1].

The paper reports secrets of why such many companies are sustainable for long years in view of KRM (Knowledge Resource Management). KUDEN（口伝） is the vital asset to keep sound management of old family enterprise other than SYAKUN(社訓), TENKUN(店訓), and OKITEGAKI(掟書) at those organizations. Specifically, KUDEN which includes a kind of management know-how is transferred to next generation by means from mouth to mouth.

Keywords: Knowledge Resource Management, Organizational Discourse, Story Telling, Dialogue, Narratives

1 INTRODUCTION

Contemporary organizations have set “SYAKUN” or “KEIEIRINEN” that are to be distributed among employees by papers. Most companies display SYAKUN in the gold-framed tablet at the guest room wall. To be sustainable in some hundreds years duration, those organizations must have some intangible assets like philosophy, mission, or vision that help manage company to alive in long life.

Japanese typical family companies who were founded 400 years ago have possessed value-driven knowledge resource as a strategic enabler. They say it is a “KUDEN”. KUDEN is not written in paper, so it is invisible. Japanese Word is “口伝”. Kanji Kuden is expressed as Mouth to Mouth communication. The traditional cake company was founded to sell sweets to Emperor in Kyoto. The company has 400 years-life and it is still very vivid in Tokyo. The company experienced many of difficult periods like change of the headquarter place from Kyoto to Tokyo about 140 years ago. At World War Ⅱ, their offices were destroyed, and during recession of 1980 decade, the company faced money problem. The CEO’s mother tells a story that some domestic battle in KYOTO burnt her family houses and shops in 16-century, and her ancestors had almost become extinct. When employees listen her narratives, they can believe the problem faced with them is not so serious. Actually, the narratives gives a power and brave to challenge to correct the family business.

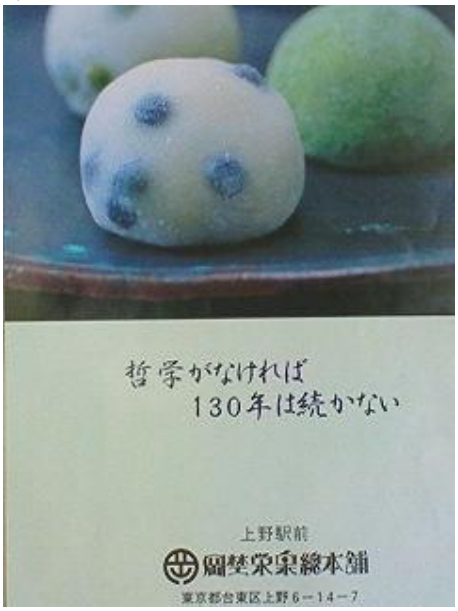
The CEO has revealed the secret of long life of the company. He follows same behavior as previous CEO, his predecessor. He repeats to talk to employees on organizational discourse, or narrative when he shares dialog time with his employee on business concepts and philosophy in a manner of a story-telling by his mouth. This way is deemed as one of unique management style in Japan. Practice of intangible knowledge by collaborative communication keeps company sustainable for long time. KUDEN is a vital knowledge to enable employee to work [2].

KUDEN is derived from Samurai and transferred to Merchant. Bushido is understood as Chivalry, and it as the Samurai’s spirit has been educated through Terakoya(寺子屋, or Jyuku(塾) by Priest-teacher’s mouth. Noblesse oblige philosophy is transferred by KUDEN which is deemed as media of “Soul of Japan”. KUDEN is a way to keep a company or team/group to survive [3].

2 KUDEN-MANAGEMENT PHILOSOPHY

For long years, KUDEN is traditionally embedded in each of organizational hierarchy through Top Management to managers and employees. At the same time, Management tells a story that explains his philosophy/concepts on managing the company. KUDEN is a kind of “theories-in-use”. When they discover serious problems through the interpretation of internal and external environments according to their knowledge, they are easy to solve the shared problems among employees. Many of KUDEN are revealed by manner of Dialogue with Top Managements and those are related to voices of customers [4]. Picture 1 shows an advertisement of Japanese old confectionery shop. It seems succeeded management philosophy is never displayed in front of customers, but their products 饅頭 are still loved by customers.

（Picture 1）



KUDEN is secretly communicated to only one son (usually eldest son). The son is almighty to produce products since he only knows how to make products.

This system is called as ISSHI SOUDEN(一子相伝). The system functions well at present and the secret of proprietary ideas are well kept. Conflicts are happened in many times regarding who get the knowhow to produce the product.

Sometimes, chief craftsman might complain of the rules. At a company producing green tea, a craftsman is permitted to know only one process, then in this system, no one is able to make complete product. The narratives so called, DENPOUKIKIGAKI(伝法聞書) is a small manual to teach the quality of product.

3 4 CIRCLES MODEL EXTENDED FROM 3 CIRCLES MODEL

3 circles models are world widely common in family business, but in Japan 4 circles model is appearing and it has demonstrated to be a good business model for keeping company long life [5]. In 3 circles model, the first circle is named Family circle. The second one is named Management one. The third one is Stockholders one.（Fig 1）

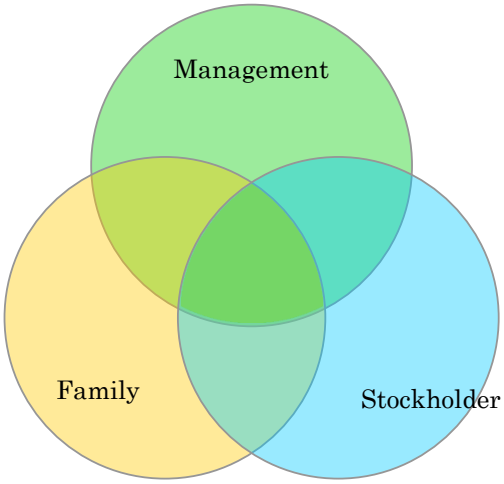


Fig.1 Family type 3 Circles Model

To add one more circle which is Associates circle to 3circles, then we may call it 4 circles model. Associates are Top management's friends who are class-mates at the university/high school, or related business friends.
Associates may give straight suggestions or frank opinions to management to correct some illegal decisions/actions before seeing actual failures.

In 4 circles model, there are two spots which belong to 3 circles. Those people who belong to two spots believe that the KUDEN is most important knowledge at the company. Without this, the company might be destroyed by some accidents like monetary problem, or a lack of corporate governance.（Fig 2）

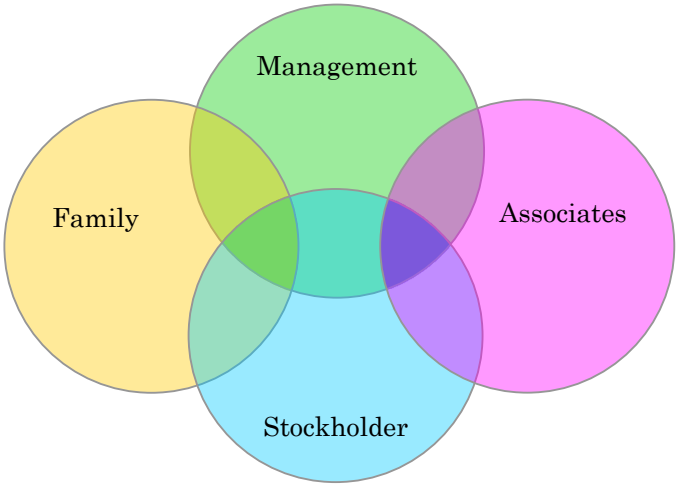


Fig.2 Associates type 4 Circles Model

Of course, typical family business is started in one circle with 3 components (Family, Stockholders and Management). Gradually, 3 circles have been built as governance model. Family circle is sometimes increasing very rapidly, then management might have cash troubles to support enough money for family-members. To keep corporate-governance has become very difficult. Most companies never went public, because family-members dislike the third-party people like bankers/security-analysts.

Recently, young Japanese entrepreneur prefers 4 circle model as a new model of sustainable organization. It is wise that the fourth circle is never superposed with family circle.

4 CONCLUSION

SHINISE(老舗) is a very old company respected and loved by Japanese customers and is not so large organization. They are not afraid of changes. They adopt new civilizations but continue to believe their culture.

Most of them are cash-rich company. Some are visionary company. Top management knows importance of Knowledge/Information and use those with their possessed KUDEN. We can say their ancients had formulated by their experiences like below. Wisdom=KUDEN+Knowledge

REFERENCES

[1] Teikoku Data Bank, Conditions for companies of more than 100 years old life, Asahi Shimbun Publication Inc.,(2009)

[2] P. H. Sullivan, Value Driven Intellectual Capital, John Wiley & Sons, Inc., (2000)

[3] F. J. Gouilart, J. N. Kelly, Transforming the Organization, McGraw-Hill, Inc., (1995)

[4] D. Grant, R. Hall, Making sense of Enterprise Resource Planning Implementations: A Narrative Analytic Approach, Journal of the Japan Management Information, Vol. 18 No 3, pp.203-220, (2009)

[5] D. Kenyon-Rouvinez, J. L. Ward, Family Business, 1st edition, Palgrave Macmillan, a division of Macmillan Publishers Limited., (2005)

Design Technique for Enhancing Software Quality and Development Efficiency

Kunihiro Yamada[1] ,Tadanori Mizuno[2]

[1] Tokai University{yamadaku@tokai.ac.jp}, [2] Shizuoka University

Taking software designing as a technical procedure will greatly contribute to the enhancement of software quality and development efficiency. To take it as a technical procedure means to get some specific processes from specifications by means of design, which also means that it is inevitable in the present technology. When dealing with software, the method of realizing a specific process covers a fairly wide scope compared with hardware, mechanisms and systems, which weakens the relationship between specifications and processes, and it eventually deteriorates the quality and aggravates the development efficiency. Hence, we present a software design technique as a single set of three items for a method of taking it as a technical procedure. This paper covers concise functional partitioning, an assessment-minded software structure and selection of the best technique, which is the most important, and it also describes the reasons for the said items.

Different types of studies and activities that deal with technical enhancement, of software are found in many fields. We see structured programming techniques for software structure [1, 2], method of quantifying software quality assessment [3], test of programming in parallel processes [4] and collaborative design test, which requires simultaneous assessment of hardware and software [5], for example. They need parallel processes of software and hardware-collaborative assessment. Problems in software do not remain within the software itself, but they greatly give effects on the process of developing hardware, mechanisms and systems. For software design and development, a number of techniques are found [6], and a variety of operations are seen as well. However, such measures have not achieved any complete solution so far.

In place of the above, our proposed software design technique offers a set of three items including concise functional partitioning, assessment-minded software structure and the best technique selected, together with the reasons for the selection. Concise functional partitioning intends to obtain a structured construction while structured designing involves an important idea in software development. Execution of optimal structuralization firstly requires a proper process of functional partitioning, and the more proper and adequate is the functional partitioning, the simpler and more streamlined the partitioning will be. Therefore, it is named "Concise Functional Partitioning", here. It is essential in the first place that the assessment-minded software structure is subjected to proper functional partitioning by means of concise functional partitioning, and a further problem is whether we can reduce the items for assessment. For the selection of the best technique and the description of its reasons, the best item is selected out of the specifications and a number of method of realization so that we can be ready to answer later confirmation.

The major task of the concise functional partitioning is how to partition and layer the functions to be achieved by the software. It aims at fewer errors made and man-hours consumed during the development, and further, it also focuses on their high extensibility and maintenability. The most focused point is that the software must have clear algorithms and applicable routines for other software to be developed later. The base of taking software design as a technical procedure is named "Selection of the best technique and the technique of describing reasons". In other words, two or more alternatives are found for embodiment of process from specifications in software design, and it is to compensate the missing processes to choose the most effective embodiment out of the alternatives.

REFERRENCES

[1] Kimiko Suzuki; "Warnier Programming Methodology", Seiri Kagaku, No. 125, pp78-89, 1973.11.

[2] CHIN-KUEI CHO; "An Introduction to Software Quality Control", John Wiley & Sons, Inc, U.S.A., 1980.

[3] Shigeru Yamada; "Way of thinking and tendency in software quality assessment", "Joho Shori" (Information Process), No. 32, pp1189-1202, 1991.11.

[4] Fumiaki Sato; "Programming test in parallel processes", "Joho Shori" (Information Process), Vol.39, No. 1, pp13-18.

[5] "Hardware and software be designed in parallel" Nikkei Electronics, No. 697, pp67-75, 1997.9

[6] PENNIE VAN TASSEL; "Practical programming technique", Nikkan Kogyo Shimbun.

Panel Discussion 2
(Chair Tomoya Kitani)

Year 2020: Green and Symbiosis Towards Post Ubiquitous Information System

Norio Shiratori[†‡] and Tadanori Mizuno*

[†]Tohoku University, Japan
[‡]Future University Hakodate, Japan
* Graduate School of Science and Technology, Shizuoka University, Japan
norio@shiratori.riec.tohoku.ac.jp

With its various boons in terms of increased efficiency, substantial economic growth, conveniences in the various spheres of life, industrial revolution also makes us forget to take care about the universe and its environment. As a consequence we can see numerous adverse effects of global warming. With the economic crisis still persistent in global market many developed countries are still reluctant to take definite steps to prevent further deterioration of world atmosphere. But it is evident that we need to take some serious action. When the world is still pondering, Japan has taken a very courageous step to control the green gas emission by a substantial margin by 2020. Though praiseworthy, but to achieve this ambitious target to $Green$, I think we need a new criterion, in addition to the rationality of modern age. Advancement in computer technology and Internet has grown beyond anybody's guess. The overwhelming presence of these two, along with many other electronic gadgets became part of our daily life, so much so that most of us, for good or for bad, cannot live without them. We can only imagine further influence of them in our life and in society. We have already witnessed that, invention of new technologies has made a huge impact in our life style and the society itself. Though it may provide various advantages and convenience, at the same time many social problems have also arisen. To solve these problems, we need an additional criterion, which we call 'α'. We believe that this new criterion along with the existing modern age rationality will lead us to the ultimate success and we can achieve a world which is both technologically efficient and environment friendly. We call this as Post Modern Society that is Green and Symbiosis Society, which will be realized around year 2020.

Ubiquitous information environment consists of Mobile Computing and Pervasive Computing, and the basis of present Modern Society. To alleviate the social problems due to the increasing gap between human and technology, an advance ubiquitous computing environment is considered, where human and information system will co-exist in cooperation with each other to the benefit of both. To achieve that we proposed a new 3^{rd} axis, comprised of human and society in addition to the existing ubiquitous computing to realize advance ubiquitous computing. By integrating these three axes, we created the Symbiotic Computing paradigm. From there a Symbiosis information system can be realized.

I think that the absence of 'α' factor mentioned above is the reason for the imbalance and Symbiotic Computing which can be considered as the 'α' is the answer. This 'α' is consisting of Green and Symbiosis. Symbiotic Computing in addition to Green will be able to create a society, where people, irrespective of their age will be able to enjoy life without the feel of inconveniences, isolation, anxiety and lack of security. The children and people of coming generation in our society can look forward to a better and Greener world and that is the main concept of Symbiotic Society that we are considering.

Consideration of industry-university cooperation in IT education

Teruhisa Ichikawa*

*Faculty of Informatics, Shizuoka University, Japan
ichikawa@inf.shizuoka.ac.jp

Abstract - 30 years ago, the discussion for IT education at the university happened at Information Processing Society of Japan. There was a remark with "The IT education of the university is not suitable for corporate needs" from the enterprise.

Afterwards, the information processing board of education started. The model curriculum was made as the university and the enterprise cooperated. The model curriculum was adopted at a lot of universities. However, a similar discussion is repeated between the industry and the academic, and the situation is not improved.

The author is in the standpoint of the enterprise 30 years ago, and exists in the standpoint of the university now. It has taken charge of the new employee education in university person's standpoint since current year. The problem in the new employee education execution is clarified. The improvement of the IT education at the university is shown.

Keywords: IT education, new employee education, educational reform, industry-university cooperation

1. INTRODUCTION

30 years ago, the discussion for IT education at the university happened at Information Processing Society of Japan. There was a remark with "The IT education of the university is not suitable for corporate needs. Therefore, the company employs good student on the deviation value and the character. And it promotes them as a specialist by the education in the company. There is no difference of the coming from faculty." from the enterprise. On the other hand, there was a remark with “There is a problem in the new employee education. The company is not considering newcomer's specialty. The newcomer is learning from elementary contents by the new employee education. It is an enterprise that prevents the difference from attaching.” from the university side committee.

Afterwards, the information processing board of education started. The model curriculum was made as the university and the enterprise cooperated[1]. Recently, an education practicing by the university came to be done. For instance, PBL(Project Based Learning) of the industry-university cooperation has been introduced[2]. However, a similar discussion is repeated between the industry and the academic, and the situation is not improved.

On the other hand, the newcomer’s adoption has decreased to each company due to the recession. It assigns to the site without doing enough group training, and it increases about the site leaving assuming the personnel training. Moreover, the new graduate employment is decreased, and it tends to increase the adoption on the way of talent who has already had technology.

The author is in the standpoint of the enterprise 30 years ago, and exists in the standpoint of the university now. It has taken charge of the new employee education in university person's standpoint since current year. The problem in the new employee education execution is clarified. The improvement of the IT education at the university is shown.

2. CURRENT STATE OF NEW EMPLOYEE EDUCATION

It investigates concerning the new employee education in Japan Industrial & Vocational Training Association, and the feature is united to monthly publication "Industrial training" and it reports. The investigation in 2007 fiscal year and the investigation in 1997 fiscal year are compared by "Feature: the realities of the new employee education and the background"[3]. Especially, the actual condition from the education before it joins a company to the following education is investigated. And, the change in the starting work consideration of the new employee is analyzed. The man power improvement education to which the interpersonal relationship ability is improved by "Promotion of the new employee who entrusts them with the future of the feature company"[4] is taken up.

Sanro research institute inc. does the case research into the new employee education. It publishes regularly in periodical "Enterprise and talent". It introduces the educational program guide and the example of some training game for the new employee in "Feature: Thus, the newcomer wants to raise it"[5].

・Chip dealings; Integration and the conflict of the team target and the personal goal are learnt from the experience.

・Interesting village; What should be of the interpersonal relationship of the problem solving in the place is learnt.

・Relay work; The problem of happening by the process of aiming at the problem achievement by the division of labor is learnt.

It introduces the following result by "Questionnaire survey concerning the new employee education in fiscal year 2005"[6].

・Especially, training to which it makes efforts

・Device that improves effect of training

・New employee education schedule in fiscal year 2005

・Education content, period, and cost per person (before join, after join, after assignment).

Japan Information Service Industry Association is introducing "Practice of an effective personnel training by the new employee education course based on ITSS" in the JISA bulletin for the new employee education related to IT[7].

The paper concerning the new employee education is few. The following papers are announced in the academic society.

Mr. Kado[8] proposed a new education method (training of pseudo OJT method by the mini project) in "Introduction training for the solution business oriented engineer promotion". It is shown that the proposal method is more effective than the current method. 83 new employees are divided by four classes. The result of the test is compared with a prior test and after test, and the educative effect is evaluated.

Table 1 Classification of new employee education

			number	hour	period
Present method	COBOL course		24	374Hr	2.5 month
	C course	studied	25	364	2.5
		not studied	20	374	2.5
New method			14	183	1.5

Ms. Omori[9] executes basic training (For eight weeks, two hours of a day every week) for 112 junior stafves (60 people are new employee) within three years joining a company, and reports on the effect.

・Acquisition of basic knowledge

・Acquisition of communications skills

・Training of autonomous thought

The problem seen through training is to miss the aspect of the system proposal and development and the evaluation that thinks about QCD(Quality/Cost/Delivery) completely. It proposes the practice of the education that the university also depressed to QCD.

Ms. Arisada[10] takes up the new employee education in the information service industry. The information service industry is a type of business that needs talent who has a special skill. However, the new employee is employed without limiting the coming from faculty. Therefore, there are lot of talents who do not have knowledge concerning the software development at all when joining a company. New employee education is executed for "Basic technological acquisition" and "Uplift of the sense of belonging to the enterprise" . The new employee education of 33 people at the IT Company is analyzed. The description concerning "Technological acquisition" and "Sense of belonging" is picked up from the report (impression) submitted every week, and new employee's change in consciousness is analyzed. As a result, a significant change in consciousness can not be seen despite the expectation of the enterprise, and it be concluded that the place of enterprise is used by new employee's various sense of values.

3. EXECUTION OF NEW EMPLOYEE EDUCATION AT UNIVERSITY

The university executed a part of the new employee education that had been done in the enterprise since April this year. It introduces the outline as follows.

There are two courses (Tokyo course for IT Company in Tokyo and Hamamatsu course for IT Company around Hamamatsu). It divided into the Tokyo office of Shizuoka University and the Hamamatsu campus of Shizuoka University. Table 2 shows the outline of the new employee education.

Table 2 Outline of new employee education

	Tokyo course	Hamamatsu course
Participant	3company 15people (8 for Business,7 for Network)	3company10people (2 for Business,8 for Embedded)
OJT	4/2～9/30(492hours)	4/2～9/30(661hours)
OffJT	4/7～5/28(210hours) (4days/week, 7.5hours/day)	4/7～7/22(204hours) (3days/week, 6hours/day)
OffJT内訳		
•Business manners	16.5hours	—
•Introduction of computer	22.5hours	18hours
•Hardware Fundamental	—	12hours
•OS Fundamental	—	12hours
•Network fundamental	7.5hours	12hours
•Database fundamental	7.5hours	18hours
•System process	22.5hours	18hours
•Control engineering	—	18hours
•Algorithm and data structure	22.5hours	18hours
•Programming	81hours (VB for Business, C for Network)	36hours (C# for Business, C for Embedded)
•System design& development practice	30hours (Business class, Network class)	42hours (Business class, Network class)
Achievement level test	1st test 4/28, 2nd test 5/28	1st test 5/27, 2nd test 7/13

Teacher's load will increase because it should execute the new employee education at the same day of the week as a usual university class and the same time zone. The course was executed by cooperating in the following lecturers around the author. It cooperated in two professors, one associate professor, and two assistants according to each one's free time. It cooperated in two visiting lecturers (1 for Tokyo and 1 for Hamamatsu) and 6 corporate lecturers (4 for Tokyo and 2 for Hamamatsu).

The author has the responsibility and the authority concerning the content of this course. The author did the composition of the course and the lecturer's selection. The cooperation between classes was able to be attempted efficiently.

The trainee was made to write the reflection seat shown in Table 3 every day. The author checked the seats and fed back to each lecturer. And opinions were exchanged if necessary. Moreover, Author attended at the presence the result of each subject, and pointed out by the viewpoint besides the charge lecturer.

Table 3 Reflection Sheet

Reflection Sheet

Date: / Name:

1. How did you feel today's course? Please rate extent to which you apply by five stages about each item (The figure of the correspondence is enclosed with ○). Please describe it if there is a special note.
   - ①Were you interested in the lecture? (Yes) 5・4・3・2・1 (No)
   - ②Was the lecture able to understand? (Yes) 5・4・3・2・1 (No)
   - ③Was the speed of the lecture early? (Yes) 5・4・3・2・1 (No)
   - ④Was the lecture difficult? (Yes) 5・4・3・2・1 (No)
   - ⑤Was the problem difficult? (Yes) 5・4・3・2・1 (No)
   - ⑥Was the composition of the lecture good? (Yes) 5・4・3・2・1 (No)
   - ⑦It is easy whether to have heard it. (Yes) 5・4・3・2・1 (No)
   - ⑧Is the overall evaluation high? (Yes) 5・4・3・2・1 (No)
2. Please describe the impressions of today's course (learnt, felt and interesting) and any question.
3. Please describe a free opinion.

The achievement level test was executed twice (the middle day and the final day of the Off-JT period). Trainee's achievement level was evaluated, and the knowledge missed by the lecture was supplemented with explaining the test.

The exercise did to the unit of the team (2～5 persons). The change seat was done at random in each subject, and the team organization was changed. "Teaching is learning of the best" was emphasized, and the motivation that guided the inexperienced person was done to the experienced person. The level difference was supplemented with teaching each other, and cooperation was able to be improved.

4. PROBLEMS AND DISCUSSION IN NEW EMPLOYEE EDCATION

The following problems became clear through the new employee education that had been executed at current year.

(1) New employee's scattering

It is common sense that electronic company employs the electronic engineering native and mechanical company employs the mechanical engineering native. However, IT Company doesn't concern new employ with specialty. Therefore, the new employees are various as shown in Table 4. The academic background is also wide with the graduate school, the faculty, Technical College, the junior college, and the special school.

It is preferable to educate from the viewpoint of technological acquisition by another class and another menu according to knowledge and the ability. However, it is difficult to organize the class according to the presence of each specialized field and the programming experience. It is economically difficult to secure more classrooms and lecturers. There is no university that organizes the class in consideration of the level of the information education in the high school. Moreover, it is preferable to work jointly with the person who has various backgrounds from the viewpoint like cooperation and the communications power, etc. New employee's scattering is not necessarily bad.

(2)Difference of corporate needs

The business system company and embed system company joined this education. Programming language is VB.NET or C# for the business system, and C for the embed system. Therefore, the programming practice and the system design & development practice executed separately for two classes.

Moreover, the control engineering was taught to eight people a company who had hope specially. The knowledge of the differentiation and physics are necessary to understand this class. However, most new employees forget it, and it was necessary to teach from the base.

Table 4 New employee's academic background

№	Class.	final academic background	Programming	1st test	2nd test	total
1	Busi.	Bachelor of Information Engineering	○C	86	70	80
2	Busi.	Bachelor of Electronic Engineering	○C,Java	69	55	64
3	Busi.	Bachelor of Material Engineering	○C,Java	72	55	66
4	Busi.	Bachelor of Mechanical Engineering	○C	58	35	50
5	Busi.	Bachelor of Business Administration	×	61	50	57
6	Busi.	Bachelor of Business Administration	×	61	45	55
7	Busi.	Bachelor of Arts	×	69	50	63
8	Busi.	Bachelor of Economics	×	53	40	48
9	Net.	Master of information media environment	○Java	97	65	86
10	Net.	Master of Bloch Science	○C++	78	85	80
11	Net.	Master of Human support science	×	86	50	73
12	Net.	Bachelor of Electronic Engineering	○C	39	50	43
13	Net.	Bachelor of management system engineering	○C++,Java	78	80	79
14	Net.	Bachelor of Biology	△Fortran	81	55	71
15	Net.	Bachelor of Regional Environment Science	×	64	45	57
			Tokyo course	65	55	65

№	Class.	final academic background	Programming	1st test	2nd test	total
1	Busi.	Master of Electronic&Informatic Engineering	○ C	80	85	83
2	Busi.	Bachelor of Information system	○ C,Java	55	85	70
3	Embed	Bachelor of Informatic	○ C,Java	70	90	80
4	Embed	Bachelor of Electronic&Informatic Engineering	○ C	60	70	65
5	Embed	Bachelor of Information Arts	○ Java	55	60	58
6	Embed	Bachelor of System Engineering	○ C,Ruby	80	90	85
7	Embed	Technical school of Information Network	○	65	65	65
8	Embed	Bachelor of System Design	×	55	70	63
9	Embed	Bachelor of Business Administration	×	50	60	55
10	Embed	High school of Electronics	×	55	60	58
			Hamamatsu course	61	71	66

(3)Short-term intensive training

The new employee education is done during half a year. Group training is done during from one to two month. It will be necessary to give the fair knowledge in this short term. When group training ends, it is assigned to each office usually. The education by OJT is executed in the assigned office. There is an office that educates based on the steady OJT plan. However, there are a lot of offices where the project person in charge leaving is promoted.

A general basic education ends at the university, and it is an ideal to do the application education and the business education in the enterprise. However, the reality is done over again from a basic education. Because they have not received systematical basic IT education. The consideration of technological acquisition of the new employee is high. Therefore, the difference of the specialized field when joining a company by the new employee education has been rapidly eased. Oppositely, the difference according to the understanding and the absorbing ability was clarified. The realities where the IT Company adopts by ability than expertise can be understood.

(4)Technological acquisition and sense of belonging

The composition and the period of Off JT were set according to the hope of the main participating company.

The Tokyo course acquired the technology by Off JT short-term(4/7～5/28) and intensively, and did OJT the

assignment ahead afterwards. Off JT trained 210 hours (4days/week, 7.5hours/day). It was admitted to go directly to the training place.

The Hamamatsu course acquired the technology by Off JT long-term(4/7 ～ 7/22) and the enterprise by OJT in parallel. Off JT trained 204 hours (3days/week, 6hours/day). To defend the working hour from going to the office to returning to the company for the Off JT period for eight hours, and to include the round trip time to the training place at working hours, and training during a day was assumed to be six hours.

There is not a difference of the achievement level test in the Tokyo course of two months and the Hamamatsu course of 3.5 months when comparing it from the viewpoint of technological acquisition as showing in Table 4 either. That is, a short-term method is more efficient. The first tension and concentration have decreased with the time passage though the attitude of attending a lecture of every day is seen.

The comparison concerning the sense of belonging is difficult at the present stage. After OJT ends, the impression to this training is asked to the executive and the person in charge of training of each company. I want to clarify the difference with this training compared with the new employee until last year.

5. CONCLUSIONS

A part of the new employee education for IT Company was executed using facilities and the teacher at the university. As a result, it was thought that the continuousness of the university's education and the enterprise's education was able to be secured. However, the reality was large the difference of new employee's knowledge. Therefore, it was necessary to teach from elementary contents of information.

New employee's technological acquisition desire was able high to be master of a basic information technology though it was training of about 200 hours.

The difference of person in question's ability came to stand out more than special different learnt at the university with the time passages. In the specialized instruction on the university that catches up by the new employee education of two or three month, the adoption policy of special disregard and the deviation value valuing of IT Company doesn't change. However, masters of IT related graduate school were top class of training from first to last, and the difference was evident.

The information education in a current faculty is a fragmentary, wide knowledge education. Therefore, even the graduate in another field can catch up easily. It was felt that it was difficult to change the recognition of the enterprise if giving priority to brushing up a basic technology and the skill, and not differentiating.

6. REFERNCES

[1] Kaminuma Y. "Information Systems Education and IS'95 Curriculum", IPSJ SIG Notes 97(27), pp.1-8(1997)

[2] Kakehi K. "Education of Informatics Major in University-J07 Curriculum Standard(<Special Feature> Trends of Informatics Education)", Journal of IPSJ, 48(11), pp.1218-1224 (2007)

[3] Japan Industrial Training Association "Special issue: The realities of new employee education and the background", Industrial Training,53(9),2007

[4] Japan Industrial Training Association "Special issue: Promotion of the new employee who entrusts them with the future of company ", Industrial Training, 55(7), 2009

[5] SANRO Research Institute "Special issue: New employee education that will start soon; newcomer wants to train it" Enterprise and talent, 40(898), 2007

[6] SANRO Research Institute "Special issue: Survey concerning new employee education of university in fiscal year 2005" Enterprise and talent, 38(864), 2005

[7] JISA "Practice of effective personnel training by new employee education course based on ITSS", JISA (90), pp.31-36 (2008)

[8] Kado T. ,Hashimoto S. "Vocational Education and Training for Novice Solution Business Oriented Engineers", IPSJ SIG Notes 2003(123), pp.41-48(2003)

[9] Ohmori K. ,Kaminuma Y. "The evaluation of software development training of the PBL style with the problem solving cycle", SEC journal 5(3), pp.204-211(2009)

[10] Arisada Y. "New employee education in information service industry" The Japan Society of Educational Sociology (48), pp.183-184 (1996)