# IWIN2009

## International Workshop on INformatics

Proceedings of
International Workshop on Informatics

September 11-17, 2009
Honolulu, Hawaii, USA

Informatics Society

Sponsored by Informatics Society

# IWIN2009

## International Workshop on INformatics

Proceedings of
International Workshop on Informatics

September 11-17, 2009
Honolulu, Hawaii, USA

Sponsored by Informatics Society

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

# Table of Contents

 **Note**: * represents excellent papers.

# A Message from the General Chair



It is my great pleasure to welcome all of you to Honolulu, Hawaii, USA, for the Third International Workshop on Informatics (IWIN 2009). 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, where a variety of interesting papers were presented. In IWIN 2009, 27 papers were accepted and 14 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. In addition, IWIN 2009 has an invited session from Dr. Norman Abramson, who is the developer of the ALOHA system.

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 Honolulu, Hawaii, USA.

Jun Munemori
General Chair
The International Workshop on Informatics

# Organizing Committee

## General Chair

Jun Munemori (Wakayama University, Japan)

## Steering Committee

Norio Shiratori (Tohoku University, Japan)
Tadanori Mizuno (Shizuoka University, Japan)
Teruo Higashino (Osaka University, Japan)
Osamu Takahashi (Future University-Hakodate, Japan)
Toru Hasegawa (KDDI R&D Laboratories, Japan)
Ken-ichi Okada (Keio University, 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)
Wang Xin (Fudan University, China)
Teruyuki Hasegawa (KDDI R&D Laboratories, Japan)
Chakraborty Debasish (Tohoku University, Japan)
Tomoya Takenaka (Shizuoka University, Japan)
Kazuya Matsudaira (Shizuoka University, Japann)
Teruhisa Ichikawa (Shizuoka University, Japan)
Kunihiro Yamada (Tokai University, Japan)
Yuko Murayama (Iwate Prefectural University, Japan)
Jun Sawamoto (Iwate Prefectural University, Japan)
Hisao Koizumi (Tokyo Denki University, Japan)
Yohtaro Miyanishi (Miyagi University, Japan)
Masahiko Ishino (Fukui University of Technology, Japan)

Yoshia Saito (Iwate Prefectural University, Japan)
Nobutsugu Fujino (Fujitsu Laboratories, Japan)
Fumiaki Sato (Toho University, Japan)
Koji Tsukada (Wakayama University, Japan)
Ryuuki Sakamoto (Wakayama 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)
Naoki Nakaumra (Tohoku University, Japan)
Norihisa Segawa (Iwate Prefectural University, Japan)
Kozo Okano (Osaka University, Japan)
Toshihiro Shikama (Fukui University of Technology, Japan)
Shohei Yokoyama (Shizuoka University, Japan)
Sanshiro Sakai (Shizuoka University, Japan)

# Session 1: Groupware
# (Chairs Tomoo Inoue and Jun Munemori)

# Synchronous and Asynchronous Discussion System Using Various Media for Improving Lecture Content

Daishiro HIRASHIMA[†], Masanori TAKAGI[††] and Yoshimi TESHIGAWARA[†]

[†]Graduate School of Engineering, Soka University, Japan
[††]Faculty of Engineering, Soka University, Japan
{dhirashi, mtakagi, teshiga}@soka.ac.jp

*Abstract* - In this paper, we focus on discussions among group members aimed at improving the content of student lecture sessions. We developed web-based system CollabSticky for improving the content of student lectures, and specifically the slide content. The system has features that allow for the submission of comments to slides synchronously and asynchronously. An evaluation of the system revealed that the users were able to make comments easily during synchronous discussion and that the system can collect a high percentage of comments which leads directly to the improvement of slide contents during asynchronous discussion. Thus, CollabSticky appears to be a promising tool not only for helping to improve presentation materials, but also to deepen the understanding of a topic among all users.

*Keywords*: Groupware, CSCW, Lecture Support, Meeting Support, Synchronous and Asynchronous Communications, Web-based Application

## 1 INTRODUCTION

Many user communication systems can now be accessed via the internet. We have been studying one such system which collects members' comments and feedback regarding meetings using presentation slides [1]. This system has also allowed us to address the problems of collecting student comments or feedback to lectures, enabling comments to be collected not only via a PC, but also on paper or by mobile phone.

In this paper, we focus on discussions among group members aimed at improving the content of student lecture sessions. The research target environment is a lecture in which students make presentations. Initially, the student is instructed to research a specific theme and then to create slides to present at the next lecture, which is followed by a Q&A session. During this session, comments for correction are provided by the teacher and questions or comments are received from the students. It is through this form of communication that the presenter and other students learn to understand better the theme of the lecture. A general model of the student lecture session is shown in Figure 1.
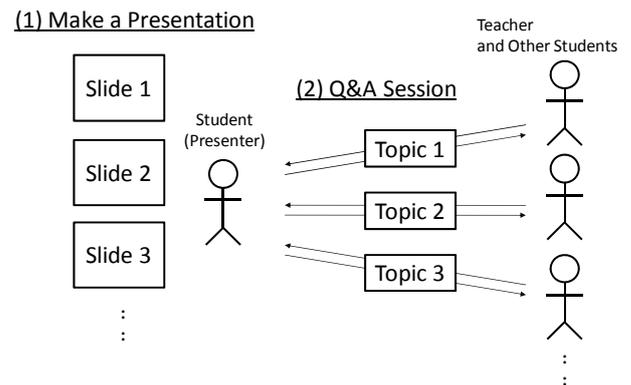


Figure 1 General Model of the Student Lecture Session

Using this model, there is a potential to understand the theme more deeply by re-examining the contents covered in the lecture, revising the slides, and discussing the topics presented among the attendees. Moreover, the revised slides themselves become good educational material. On the basis of this model, here we propose a system for collecting comments as feedback of student lectures and to facilitate asynchronous discussions with the aim of improving the lecture content presented on slides.

## 2 SYSTEM REQUIREMENTS

CollabTest [2] is an existing system designed to improve quiz content during lectures (synchronous feedback) and after lectures (asynchronous feedback). In this system, the student makes a quiz together with explanations about a specific learning topic assigned by the teacher. As the student likely makes some mistakes or presents information confusingly in the quiz, other students provide comments as feedback to revise the content. Asynchronous discussions through this system enable high-quality quiz content to be finalized. The concept of CollabTest for improving quiz content is shown in Figure 2, and a comparison of the features of this system and the proposed system is shown in Table 1.
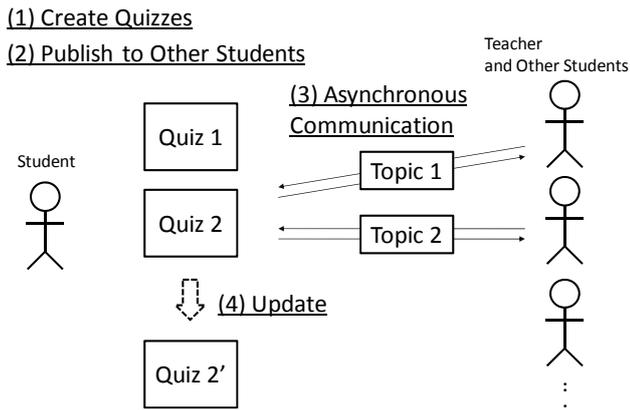
Figure 2 Concept of the CollabTest System for Improving Quiz Content

Table 1 Comparison of CollabTest and the Proposed System

|  | CollabTest | Proposed System |
|---|---|---|
| Communication Environment in Two Phases | Synchronous and Non-Face-to-Face | Synchronous and Face-to-Face |
|  | Asynchronous and Non-Face-to-Face | Asynchronous and Non-Face-to-Face |
| Target Data | Text data (Quiz) | Image data (Slides) |

## 2.1 Target Environment and Comment Posting Model

CollabTest and the proposed system are designed for use in the same environment, where the communication environment moves from an initial synchronous phase (during the lecture) to an asynchronous phase (after the lecture).

Target data in the case of CollabTest starts from a Synchronous and Non-Face-to-Face environment since the students create quizzes by themselves. In contrast, in the case of the proposed system, students send opinions or questions synchronously during the student lecture and Q&A session. This requires a comment posting feature as well as a slide view feature to be used synchronously. The comment posting model of CollabTest and the proposed system is shown in Figure 3.
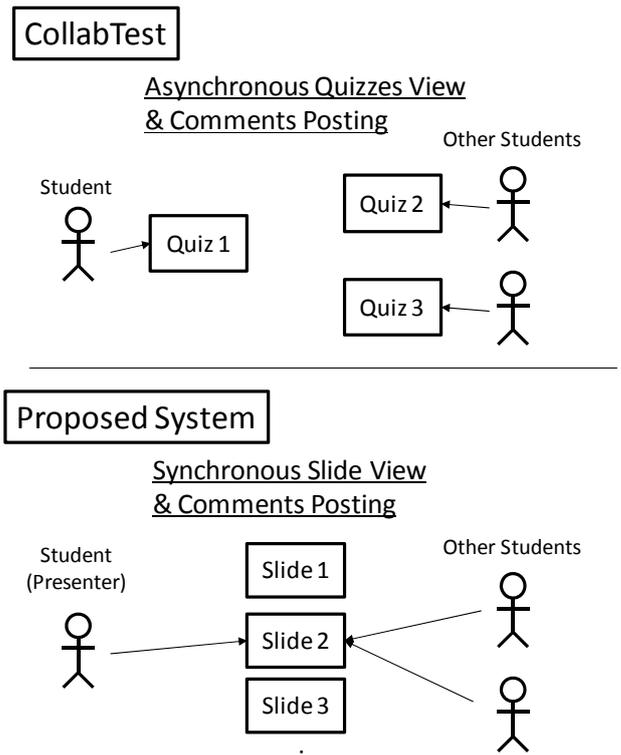


Figure 3 Comment Posting Model of CollabTest and the Proposed System

## 2.2 Target Data and Referring Method

In terms of data features, the target data for improvement with CollabTest is quiz data which is text-based, whereas it is slide data which is image-based with the proposed system. CollabTest is able to improve quiz content via text chat communications. However, it is not sufficient for the proposed system to use text chat communications to improve slide content as there are many slides and many topics on the slides to be discussed, which means the target topics tend to be ambiguous. Thus, it is necessary to have a feature which highlights a specific point or section of the slide on the x and y axes and registers comments in that specific location, thus minimizing ambiguity. Figure 4 shows the comment reference model of CollabTest and the proposed system.
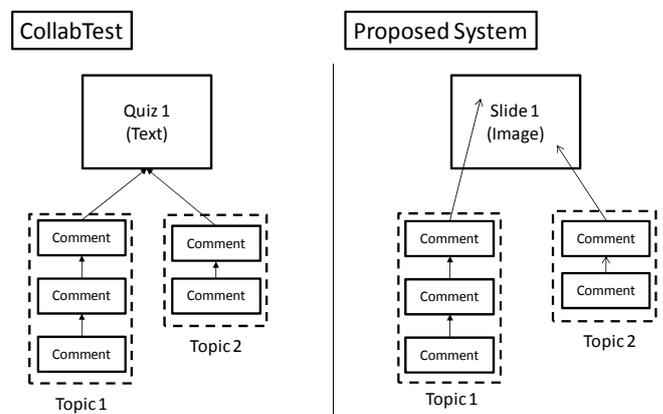


Figure 4 Comment Reference Model of CollabTest and the Proposed System

## 2.3 Discussion Topic Transition from Synchronous to Asynchronous Environment

The comments submitted by students using the proposed web-based system, which we call "CollabSticky", become discussion topics in the asynchronous environment after the lecture session. This system makes a chat thread for every comment, which we call a discussion thread. Figure 5 shows the transition between comments submitted synchronously during the lecture and the discussion thread used after the lecture.
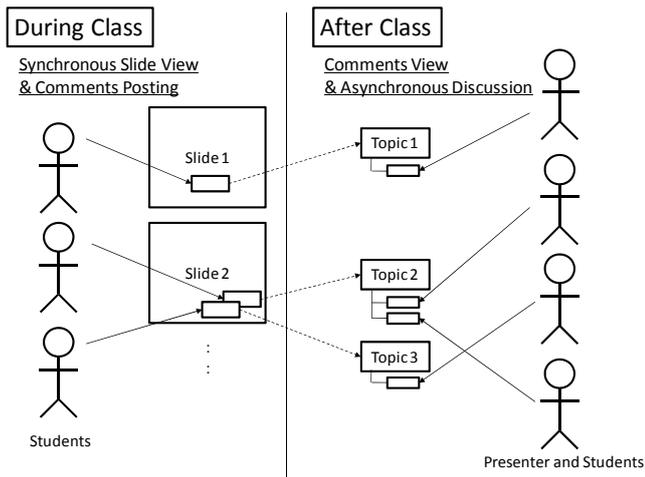


Figure 5 Transition between Submitted Comments during the Lecture and Discussion Threads after the Lecture Using the Proposed CollabSticky System

When someone submits a comment during asynchronous discussion after the lecture, CollabSticky sends an email about the comment to other students to facilitate a quick response.

## 2.4 Proposed System Utilizes Various Media

We have been utilizing the proposed CollabSticky system in our investigations into how best to collect and encourage comments from students during lectures. In this research, CollabSticky is designed to collect a large number of comments as feedback not only via a PC, but also on paper or by mobile phone. This means that the target environments of this research are not only more specialized classrooms having PCs installed for each student to use, but also ordinary class rooms without PCs.

## 3 THE PROPOSED COLLABSTICKY SYSTEM

CollabSticky supports PowerPoint slide files. The student presenter uploads his/her PowerPoint file to the system by using a web browser and the server side program then converts the file to jpeg image files. The student then selects the name of the PowerPoint file from the list displayed and CollabSticky displays the first slide in the file. Figure 6 is a screen shot of CollabSticky where users can change to next or previous slides by clicking the right or left arrow in the

browser, using a mouse wheel or keyboard shortcuts. In addition, users can see thumbnail images of all the slides in the presentation file. When users click on one of the slide thumbnails, CollabSticky displays the selected slide.
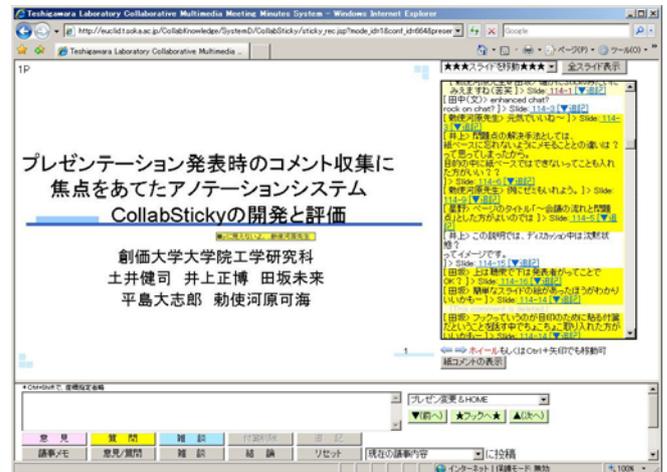


Figure 6 CollabSticky

One of the useful features of this system is rapid slide change. As CollabSticky stores image cache files before or after the current slide and uses Ajax, slides can be changed very quickly.

## 3.1 Synchronous Slide Presentation and Comment Submission

We use the "Slide View" function for synchronous discussion during the lecture session (Fig. 8, left). The user who clicks the "Make a Presentation" button is recognized as the parent user, and the system then shows the first slide in full screen mode.

Left of this button, there is the "Synchronous Slide" button. CollabSticky recognizes the user who pushes this button as a child user, whose slide changes synchronously and automatically with the parent slide.

When the child user clicks on a particular part of a slide, the textbox below is highlighted and the user can then input text comments. The comment submitted via this textbox is displayed at exactly the location on the slide the child user selected. This is like putting a sticky vote on the slide. CollabSticky recognizes this sticky vote as a parent comment, displaying the same slide to all students but not the presenter at the same time. Therefore, CollabSticky shows not only the same slide synchronously, but also the same sticky votes in real time.

When another student clicks the parent comment, the mode changes to "Adding Comment Mode" and a textbox below is highlighted. A submitted comment in this mode is added to the parent comment which the user just clicked. CollabSticky recognizes this submitted comment as a child comment. When a user mouses over the comment, the parent comment is extended with children comments shown below. In order to view the slides easily, only the parent comment is shown and is translucent. Figure 7 shows a child comment viewed with mouse over.

Figure 7 Child Comment Viewed with Mouse Over

These system features enable comments to be submitted synchronously regarding a specific part of a slide on the x and y axes and for students to communicate with each other in real time.

## 3.2 Extending Synchronous Discussion to Asynchronous Discussion

In the synchronous discussion, CollabSticky recognizes each submitted parent or child comment as one discussion thread. In the asynchronous discussion after the lecture, users can view all the discussion threads in "Sticky View" (Fig. 8, right).

In "Sticky View", every discussion thread referring to a specific slide is displayed. When a user clicks the "Submit Comment Here" link below the discussion thread, a textbox and a submit button appear, allowing the user to submit a new child comment. Figure 9 shows the submission of a new child comment in the asynchronous discussion environment.

After the lecture, the student presenter can respond to all the submitted comments. However, as we found that submitted responses were not being viewed quickly by the other students, we developed a feature to inform the students by email about the comment and response submissions. We speculated that this feature would increase the speed of communication.

We also speculated that it is efficient for discussion and feedback to involve not only the student presenter and students who submitted comments, but also other group members. Therefore, those members to be informed by email are chosen by the student submitting a comment. This feature should help the members considered most appropriate for a particular topic to join the discussion.

## 3.3 Collecting Comments Using Various Media

It is necessary to collect comments from many students efficiently. However, in actual lectures, it is not always possible for all students to use their own PC. Additionally, not all students may have the ICT skills needed to input all the comments they want to submit. Therefore, we developed a feature to collect comments on paper or by mobile phone.

In the case of collecting comments by mobile phone, CollabSticky collects them via email. Figure 10 shows how comments are input using a mobile phone. Students input the slide number in the title field and the comment in the body of an email. CollabSticky receives and analyzes the email and stores the data in the database.
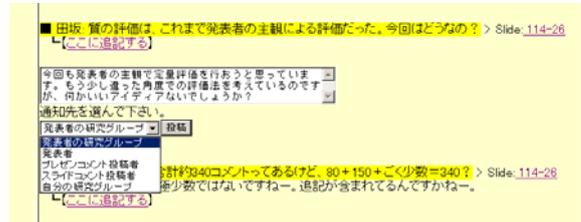


Figure 9 Submitting a New Child Comment in Asynchronous Discussion



Figure 8 "Slide View" for Synchronous Discussion (left) and "Sticky View" for Asynchronous Environment (right)

Figure 10 Comment Input Screen for Mobile Phone

Figure 11 shows the form used to collect comments on paper, where comments can be clearly and specifically marked relating to the content shown for three slides. CollabSticky converts all of the collected papers to image data and stores them in the database. These are displayed as thumbnails and the original image is shown when the user clicks the thumbnail. The "Slide View" feature has a link to this thumbnails page.
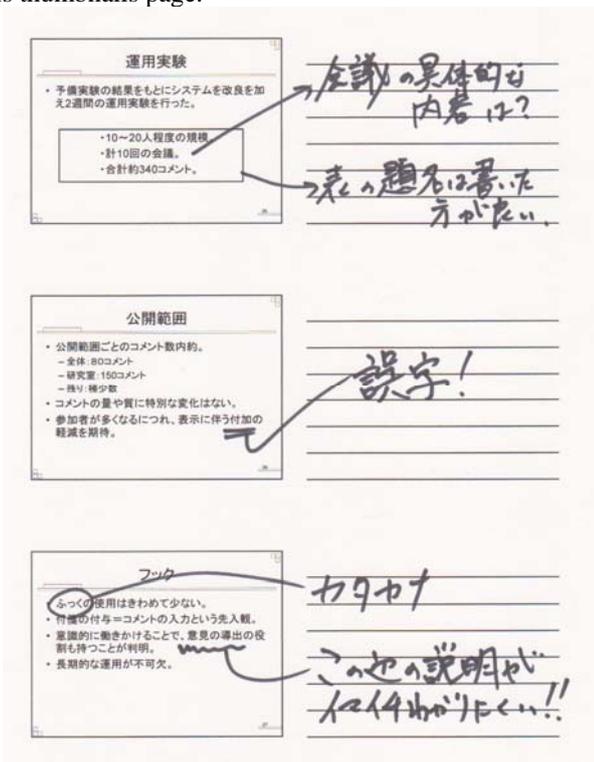


Figure 11 Example of Paper-based Comments

## 4 EXPERIMENTS AND EVALUATION

We confirmed that the features described in section 3 were normally executed and that the system operated without problem. In the next section, we describe system evaluation in the synchronous and asynchronous phases of discussion, and involving the collection of comments by mobile phone or on paper.

### 4.1 Synchronous Comment Submission

One of the important objectives of using CollabSticky is to collect large numbers of comments in order to improve the lecture content most effectively via discussion after the lecture. However, during the lecture itself, it is not easy both to understand the slide contents and submit comments by using CollabSticky; the system is useful only if the effort required to enter comments using the system is not prohibitive and allows adequate numbers of comments to be submitted.

Therefore, we devised experiments in our laboratory to ascertain how many comments might be submitted during actual student lecture sessions. A group consisting of one presenter and four students participated in the experiment. The students have ICT skills. In order to create an environment in which the students would acquire new knowledge, we created slides about four topics which were far from the students' research fields. After each of the lectures, there was time for Q&A. During both the lecture time and the Q&A session, the four students submitted comments by using CollabSticky via a web browser. Tables 2 and 3 show the duration of the lecture sessions and the number of parent comments received, respectively.

Table 2 Duration of Lecture Sessions (in minutes)

|              | Lecture | Q&A   | Total |
|--------------|---------|-------|-------|
| Experiment 1 | 20:21   | 12:02 | 32:23 |
| Experiment 2 | 8:00    | 17:05 | 25:05 |
| Experiment 3 | 22:32   | 13:46 | 36:18 |
| Experiment 4 | 9:52    | 14:49 | 24:41 |

Table 3 Number of Parent Comments Received During the Lecture Sessions

|              | Lecture | Q&A | Total |
|--------------|---------|-----|-------|
| Experiment 1 | 7       | 9   | 16    |
| Experiment 2 | 8       | 7   | 15    |
| Experiment 3 | 15      | 13  | 28    |
| Experiment 4 | 13      | 6   | 19    |

We judged from the results that sufficient numbers of comments could be collected using CollabSticky. In the case of experiments 3 and 4, the number of comments received during the lecture itself exceeded that received during the Q&A session. Thus, it appears that the effort required to use the system was not prohibitive for the students to make adequate numbers of comments.

In addition, we categorized the submitted comments into four groups, as shown in Table 4. Table 5 shows the number of comments by category.

Table 4 Comment Categories

| Opinion    | Comment in opposition to slide content, or indication of presenter's mistake |
|------------|------------------------------------------------------------------------------|
| Question   | Comment requesting additional information about slide content to aid greater understanding |
| Impression | Comment on slide without requesting action from the presenter |
| Other      | Chat or other comment which does not fall into the other three categories |

Table 5 Number of Comments by Category

|  | Opinion | Question | Impression | Other | Total |
|---|---|---|---|---|---|
| Experiment 1 | 3 | 13 | 0 | 0 | 16 |
| Experiment 2 | 0 | 14 | 0 | 1 | 15 |
| Experiment 3 | 8 | 16 | 3 | 1 | 28 |
| Experiment 4 | 8 | 8 | 2 | 1 | 19 |
| Total | 19 | 51 | 5 | 3 | 78 |

As the presenter should respond to opinion comments and question comments, these two category comments show high potential for discussion among the presenter and other students. Therefore, the experimental results suggest that CollabSticky can collect up to 90% of comments which lead to asynchronous discussion.

## 4.2    Asynchronous Discussion and Improving Slide Content

We next conducted an experiment to investigate how much asynchronous discussion and revision of slides might result form using CollabSticky. The participants were 5 graduate students and a teacher.

After a graduate student in the role of presenter uploaded a file to CollabSticky, the other students and a teacher submitted comments on the slides as described in section 3. Because the student presenter had to give a presentation in a plenary meeting in the laboratory in the near future, we compared the uploaded slides to the revised slides created after asynchronous discussion by using CollabSticky, and we investigated the number of revisions made.

Table 6 shows the number of submitted comments by category.

Table 6 Number of Submitted Comments by Category

|  | Opinion | Question | Impression | Other | Total |
|---|---|---|---|---|---|
| Experiment | 19 | 6 | 5 | 0 | 30 |

There were 25 submitted comments categorized as Opinion and Question comments which demanded responses from the student presenter. Of these 25 comments, 18 received a response comments from the presenter. We considered these 18 comments as 18 asynchronous discussion threads and checked whether CollabSticky sent an email to students when each comment and response was submitted.

Ultimately, the presenter felt that 14 comments were helpful to improving his presentation and he revised the slides accordingly. The presenter remarked that this system was very useful to help him improve his presentation, and specifically the content of his slides.

These results indicate that CollabSticky was effective for improving content during asynchronous discussion. However, the presenter did not respond to 7 submitted comments. At this point, CollabSticky does not have a function to follow the status of each discussion thread, so we were unable to determine easily whether the comments were considered irrelevant by the presenter or were forgotten. CollabTest has a navigation feature for discussion threads that indicates which comments were fund to be

helpful to improve the quizzes created and it performs well [2]. In future, we plan to add such a feature to the CollabSticky system

## 4.3    Collecting Comments By Mobile Phone or On Paper

As described in section 3.3, CollabSticky collects comments submitted by mobile phones or on paper. In an experiment to determine the effectiveness of using these two features, we found that no asynchronous discussion threads were generated using these features. This appears to be due to the fact that the users could make insufficient reference to specific slide content; users found it difficult to refer to specific points on a slide, which created ambiguity in system use.

As described in section 2.2, it was necessary that asynchronous discussions made reference to a specific point on the x and y axes in a slide in order to avoid ambiguity of comments. Therefore, in the future, we will attempt to apply the same method in the case of submitting comments by mobile phone or on paper.

## 5   CONCLUSION

This paper proposed the CollabSticky system for improving the content of student lectures, and specifically the slide content. The system has features that allow for the submission of comments about slides which are changed synchronously with the presenter's slides, the submission of comments that refer to specific locations on the slides, the transitioning of collected comments in the synchronous environment to the asynchronous environment, and communications that promote improvement of slide content itself.

An evaluation of the system revealed that the users were able to make comments easily during synchronous discussion and that the system can collect a high percentage of comments which leads directly to the improvement of slide contents during asynchronous discussion. Thus, CollabSticky appears to be a promising tool not only for helping to improve presentation materials, but also to deepen the understanding of a topic among all users.

We plan to improve CollabSticky by including a navigation tool that indicates the status of discussion threads and to improve the reference to specific points in slides when using the features of comment submission by mobile phone or on paper. We will use CollabSticky in real lectures in order to implement these system changes effectively and will evaluate it in terms of learning effects.

# REFERENCES

[1] Kenji DOI, Daishiro HIRASHIMA, Masanori TAKAGI, Masamitsu MOCHIZUKI, and Yoshimi TESHIGAWARA, Development of a Lecture Supporting System to Collect Comments Using Various Media, Proceedings of the IASTED International Conference Web-based Education (WBE 2009), pp.294-300, 2009.3

[2] Daisuke HOSHINO, Masanori TAKAGI, Noriko MINAMI, Yoshimi TESHIGAWARA: Navigation Function of Group Review for Promoting Collaborative Improvement of Quizzes Created by Students, Proceedings of the IASTED International Conference Web-based Education (WBE 2009), pp.353-359, 2009.3

# Proposal of an Idea Generation Support System Using Digital Photographs with Position Information

Takahiro Matsui*, Junko Itou **, and Jun Munemori **

*Graduate School of Systems Engineering, Wakayama University,Japan
**Faculty of Systems Engineering, Wakayama University, Japan
{s105050, Itou, munemori}@ sys.wakayama-u.ac.jp

*Abstract* - Digital cameras have come to be widely used. For example, we use digital photographs for blog articles, travel diaries, and to record memories. There are numerous opportunities to take digital photographs with several people. Photographs are used as entertainment or as field work data. We have developed GUNGEN-Photo. GUNGEN-Photo is an idea generation support system using digital photographs with position information. The system is performed with several people who use group operation of idea generation. We aim at new idea generation using photographs. We experimented with the system and evaluated it.

*Keywords*: Digital photograph, Position information, KJ method, Table top interface.

## 1   INTRODUCTION

Idea generation methods are of great importance. There are many idea generation methods [1], [2]. In addition, many systems have been developed to support them [3], [4], including those that use a Table Top interface [3], [5].

Digital cameras have come to be widely used. The acquisition of digital photography has become easy. We take pictures frequently when we discover interesting things during travel. Dispatching various information using digital photographs is enabled. For example, we use digital photographs in blog articles and SNS [6]. There are opportunities to talk about digital photography with various people [7]. Photographs are used as entertainment or field work data. It has become easy to dispatch information using photographs.

We have developed a system, "GUNGEN-Photo", which is an idea generation support system using digital photographs with position information by several people talking about photographs while meeting. This system seems to be one of the idea generation methods using photographs. We carried out an application experiment of idea generation using several people with this system. We tested taking a trip by this experiment. We began by writing each impression and an opinion in a caption for the photograph, which had been taken in town. We found new discoveries such as memories or the problems of the trip with them. Such discoveries may produce good ideas.

Chapter 2 explains the KJ method. In Chapter 3 we explain GUNGEN-Photo. Chapter 4 shows the experiment that uses this system. Chapter 5 describes the experiment results. Chapter 6 is a discussion, and chapter 7 is the summary.

## 2   THE KJ METHOD FOR PHOTOGRAPHS

This system is influenced by Kawakita's generation system, called the KJ method [2]. The KJ method was developed by Jiro Kawakita. The KJ method is one of the most famous idea generation methods in Japan. The KJ method is known as a method for establishing an orderly system from a chaotic mass of information. The method was originally developed for anthropological field-work to delineate relations that would lead to findings from gathered and stored data. The KJ method as applied to technical innovations involves the systematization and covering of brainstorming. The feature of the KJ method is cooperative work toward innovation. The KJ method is suitable for the tabletop interface [8]. The four steps of the KJ method are as follows.

(1) Label making (brainstorming step)

In the first step, participants write their ideas freely on tags (i.e. Post-it notes) according to the theme of the meeting. Then the tags are put on a table. The tags are referred to as labels. When participants write ideas, they should not criticize other's ideas, and they must propose their ideas without hesitation. This step corresponds to brainstorming, and deals with photographs and article making. (We call photographs PhotoLabel and articles CommentLabel here; both are referred to as *labels* in this system.) The PhotoLabel is a photograph which was taken at a particular location. The CommentLabel is a comment about photographs.
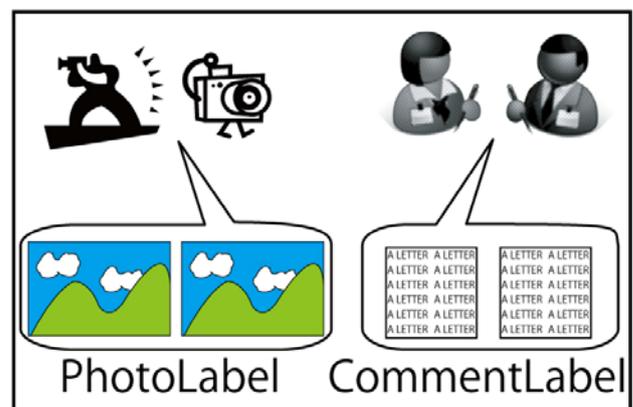


Figure 1: The label making step

(2) Grouping ideas

In the second step, participants examine their labels and group them into groups through discussion. The criterion for this grouping is not the category of ideas but their intuitive similarity. Each group is called an island and given a representative title (naming). The island is a technical term in the KJ method, and a set of similar ideas. This system performs grouping with both PhotoLabel and CommentLabel.
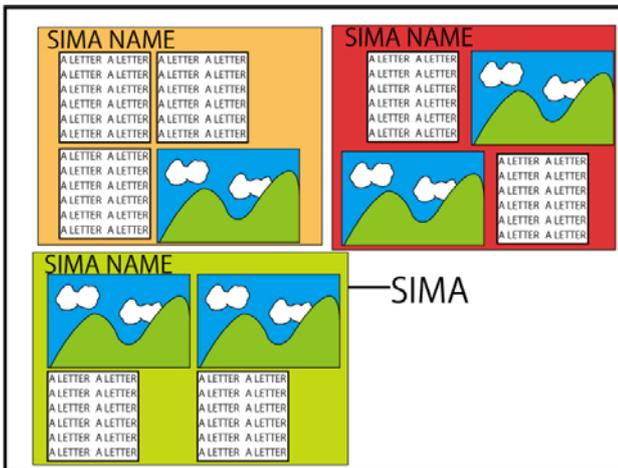

Figure2: The idea Grouping step

(3) Structuring groups

In the third step, participants look for an arrangement that express the mutual relations of the representative titles spatially. Then they connect the related representative titles together in a certain line. This system does not currently support this step.

(4) Writing a composition

In the last step, participants write a summary composition. They should not express their opinions but should write it based on the data itself. Our system is a process to make new ideas from PhotoLabel and CommentLabel.

These four steps are the working process of the traditional KJ method. The KJ method involves a great deal of communication (face-to-face meeting). Participants can carry out the KJ method by labels and a large sheet of paper (i.e. Japanese B0 size paper). A participant can easily point to an object (a label or an island) on the paper.

## 3　GUNGEN-PHOTO

We have developed GUNGEN-Photo. This system is an idea generation support system using digital photographs with position information. This system is performed with several people who use group operation of the idea generation. This system uses a DiamondTouch-Table. We can realize face-to-face communication by direct manipulation.

### 3.1　Composition of system

The system consists of the following hardware and software.
(1)　Hardware
    DiamondTouch-Table（65cm×87cm）[9]
    CANON DATA PROJECTER X700
    DELL (Intel (R) Core 2 Quad CPU Q9450 (2,66GHz))
(2)　Software
    DiamondTouchSDK2.1 (MERL)
    Microsoft Windows XP Professional version 2002
    JDK1.6.0-07(Sun Microsystems)
    Eclipse SDK ver.3.4.1 (2000 lines program)

### 3.2　Operation of system

We shall now describe the main operation of the system.
(1) Initial screen
Figure 3 shows the initial screen of the system. Five kinds of buttons are shown in the tool window (from the perspective of the user, it is on the left of the screen.). If we click the comment card generation button (b), a label is generated. When the island making button (c) is clicked and it is in the ON state, the system displays the range of the island by appointing the range of the island that we generate with both hands. When the label or the islands delete button (d) is clicked and it is in the ON state, it can be deleted. When the island resize button (e) is clicked and it is in the ON state, the size of the island is changed by the hand pointing. If we click the position lining up button (f), and touch a PhotoLabel, we can arrange scattered PhotoLabels in a line. (c)~(f) buttons indicate change in ON and OFF.
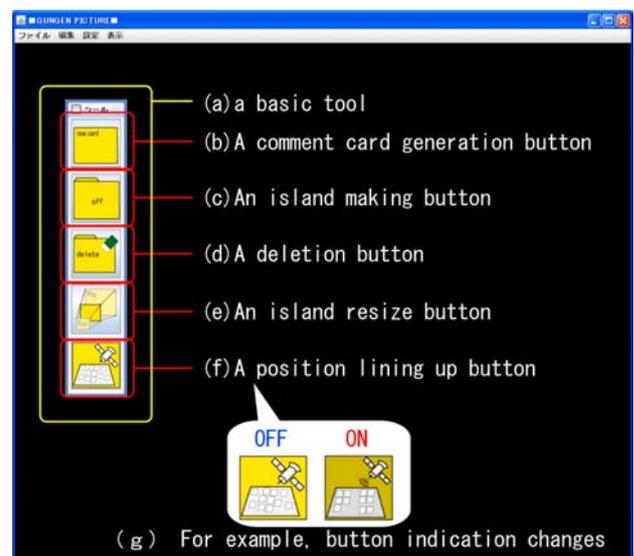

Figure 3: Initial screen

(2) Starting brainstorming
Figure 4 shows an example of the initial screen of brainstorming. After grouping PhotoLabels, we generate CommentLabels, and describe participant's ideas on the labels. If we click the CommentLabels and islands, we can input the characters and set color islands. The system has no flow control. We can operate each object simultaneously.

Figure 5 shows an execution screen of the system. Some CommentLabels are added to islands.
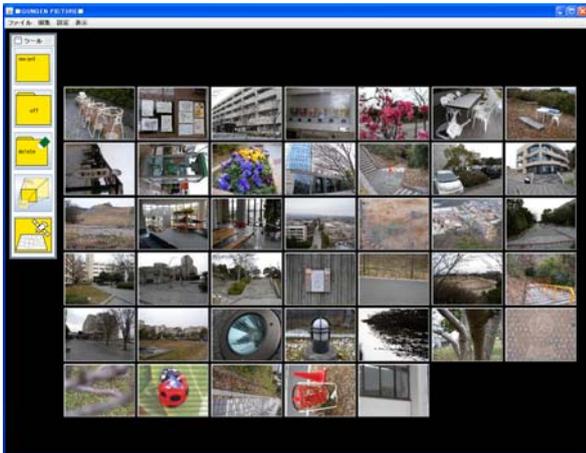

Figure 4: An example of the initial screen of brainstorming
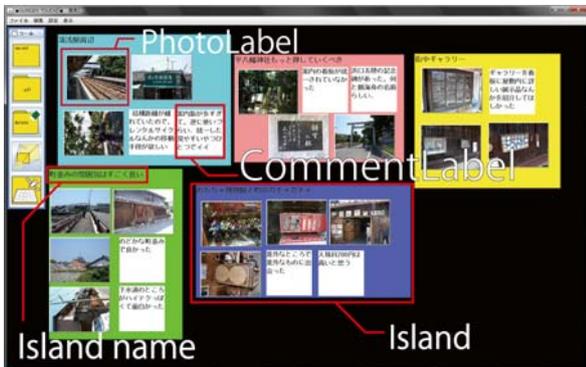

Figure 5: System execution screen

(3) Functions of system

We shall next explain functions except for the basic brainstorming tool. There is an expansion function for photographs. The photographs on the system are displayed as thumbnail images. Their sizes are 160pixel x 120pixel. This system makes the thumbnail image automatically when photographs are inputted. However, when a user needs to view a photograph in detail while working, the enlargement function can be used. If he or she touches a photograph when all buttons of the basic tool are off, the system enlarges the photograph to 480pixel × 360pixel. The size of the thumbnail images is 4 to 3. We decided the ratio from the aspect ratio of a digital photograph which we photographed with a general digital camera.

It is not easy for us to distinguish between photographs displayed in large quantities. We made the simple identification function of the photograph to assist the work of distinguishing between photographs. This is the lining up function of the photographs, which uses position information. An explanation of it is as follows. Figures 6 and 7 show examples of photographs aligned with this function. We choose one photograph from photographs displayed separately when the position lining up button is ON. As a result, a dialogue appears. We input distance information into a displayed dialogue in Figure 6. We can freely input the numerical value in a metric unit. Lining up is started when we push the start button of the dialogue. We line them up again with other photographs fulfilling a

condition of the numerical value input mainly on the position information of the photograph that we chose. Figure 7 expresses the results of lining them up. There is no change in the photographs that do not fit that condition. When lining up is over, the position lining up button returns to OFF.
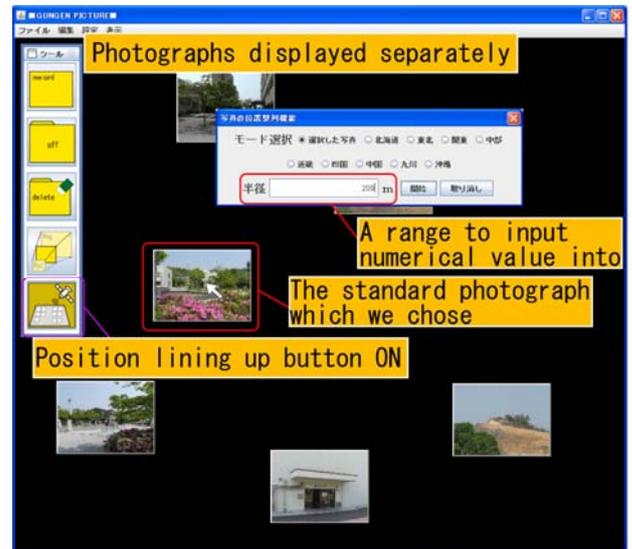

Figure 6: An example before using the position lining up function
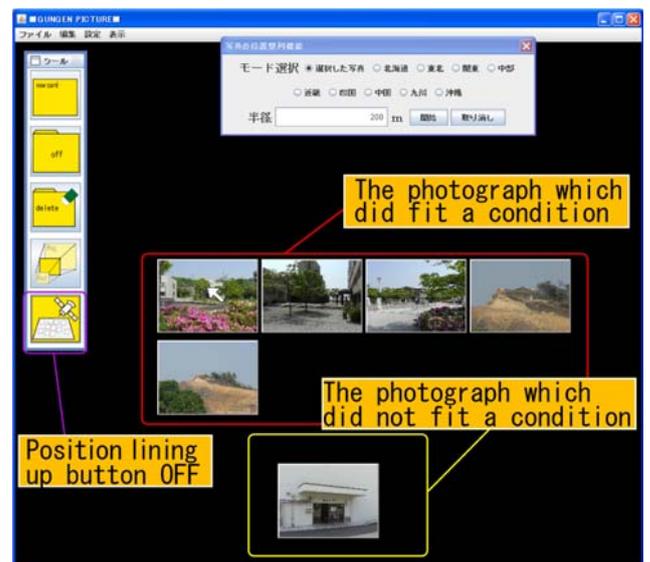

Figure 7: After using the position lining up function

## 4 EXPERIMENT

We have carried out 2 experiments with GUNGEN-Photo. After the experiments subjects completed a questionnaire. The first experiment checked the effectiveness of identification by position information of digital photographs. In addition, we checked the grouping work of photographs with several people. One group consisted of four participants. Subjects were asked to take 20 pictures in Wakayama University. The camera the subjects used was a NIKON COOLPIX P6000. This camera can add position information to photographs automatically. We had subjects take photographs, then arrange them on GUNGEN-Photo.

40 photographs were used in this experiment, ten from each subject. Photographs were chosen at random. The number of participants was 16 and they were divided into four groups (Group A ~ Group D). All participants were students of Wakayama University and all belonged to the faculty of system engineering. The subjects perform grouping of all PhotoLabels by similarity. Once they give all islands names, the experiment is finished.

The second experiment was carried out to check the effect of the idea generation support system using digital photographs with position information by various people. The second experiment was performed by five groups (Group A ~ Group E). A group consisted of three people. The number of participants was 15. All participants were students of Wakayama University and belonged to the faculty of system engineering. We asked subjects to take pictures in the town of Yuasa in Wakayama Prefecture. We had subjects take photographs by group unit in experiment 2. We did not appoint a particular number of photographs. 40 photographs were used in this experiment, which were chosen freely by the subjects.. They carried out idea generation using digital photographs with position information of Yuasa town with several people on GUNGEN-Photo. At first we had subjects group the PhotoLabels. We call this work procedure 1. After grouping PhotoLabel, subjects made CommentLabels. We had subjects group both PhotoLabels and CommentLabels once again. Once the naming of the group was finished again, the experiment was finished. We call this work procedure 2.

## 5    EXPERIMENTAL RESULTS

### 5.1    Experiment 1 results

The results of experiment 1 are shown below. Table 1 shows the results of each experiment. The units of working time in Table 1 are minutes and seconds. A~B of Table 1 are group names. "AV" under the group name is the average of each parameter. The results of the questionnaire are shown in Table 2 and Figure 8 (five point scale evaluation)."5" is the highest score and "1" is the lowest. Each of the people in Figures 8 and 9 are the same. Table 2 is average of questionnaire on the position lining up function of the photographs. Figure 8 is the remark counts according to the subject of each group. The count of the remarks affects work in experimenting. The content of the remarks is not shown. Figure 9 shows questionnaire results concerning the merits of performing with various people according to the subject.

Table 1: The results of each group

|   | Working time | Number of remarks | Position lining up function | Image enlargement function | Number of islands |
|---|---|---|---|---|---|
| A | 21m18s | 299times | 4times | 13times | 6units |
| B | 47m22s | 242times | 3times | 12times | 7units |
| C | 33m22s | 235times | 3times | 9times | 11units |
| D | 25m20s | 598times | 5times | 19times | 6units |
| AV | 31m51s | 343.5times | 4.3times | 13.3times | 7.5units |

Table 2: The experiment 1 questionnaire

| About the position lining up function the photograph | |
|---|---|
| 1: Did you use the arranging photographs function? | 4.4 |
| 2: Did you think that this function was useful? | 4.2 |
| 3: Was this function a trigger to arrange photographs | 4.0 |
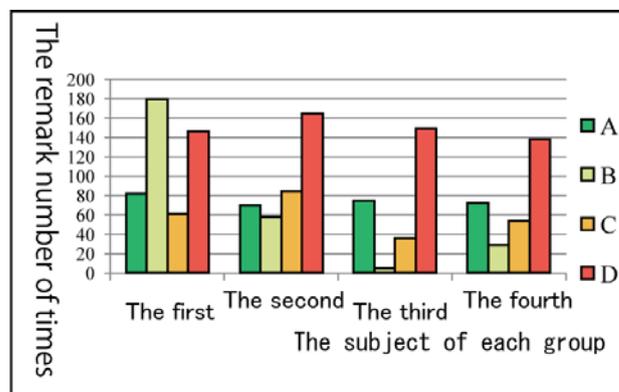


Figure 8: The remark counts according to the subject of each group
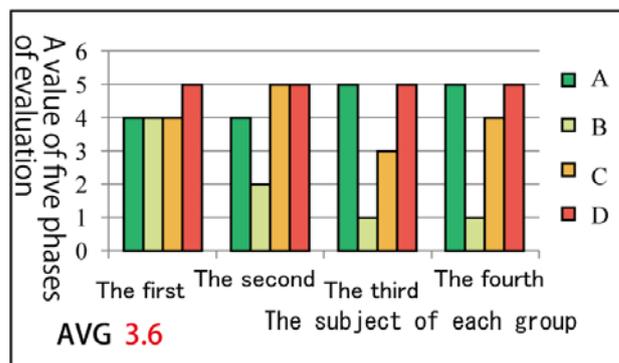


Figure 9: A questionnaire for the merit to perform with plural people

### 5.2  Experiment 2 results

The results of experiment 2 are shown below. Table 3 shows the results of experiment 2. "m" and "s" indicate minutes and seconds. Tables 4 and 5 show the average number of remarks of subjects during each work. We divided remark contents into 4 unlike experiment 1 (The contents of the remark were opinions, agreement, questions and answers). Tables 6 and 7 are the questionnaire results after finishing experiment 2. In Table 6,"5" is the highest score and "1" is the lowest. Table 7 shows description type questionnaire results.

Table 3: Results of experiment 2

| | |
|---|---|
| The average number of islands when grouping PhotoLabels | 8.2units |
| The average number of islands when making CommentLabels | 8.2units |
| The numbers of CommentLabels | 12.6units |
| The average working time of grouping PhotoLabels | 18m19s |
| The average working time of making CommentLabels | 27m16s |

The number of islands in procedure 2 was the same as procedure 1. The number of comments was 0 to 4.

Table 4: The average number of the counts according to the type of subjects' remarks in procedure 1

| | |
|---|---|
| Opinion | 97.4 |
| Agreement | 47.4 |
| Question | 22.2 |
| Answer | 15.2 |
| Total | 182.2 |

Table 5: The average number of the counts according to the type of subjects' remarks in procedure 2

| | |
|---|---|
| Opinion | 180.4 |
| Agreement | 103 |
| Question | 23.6 |
| Answer | 20 |
| Total | 327 |

The number of remarks expressing "opinion" and "agreement" increased when making a CommentLabel. In contrast, "question" and "answer" hardly changed. Comparing procedures 1 and 2, the ratios of the agreement remark counts among remarks increased to 0.31 from 0.23.

Table 6: Questionnaire results of experiment 2

| | |
|---|---|
| Do you think that there is a merit in operating with various people | 4.7 |
| Was the CommentLabel making function effective? | 4.4 |

Table 7: Description type questionnaire results of experiment 2

| |
|---|
| The conversation becomes changed when we look back on comments while viewing photographs. |
| Viewing photographs stimulates lively Conversations and I can remember things I had forgotten. |
| We were able to think of at what point there was a problem while looking around the whole. |
| I want you to make the sensitivity of DT better. |

## 5.3 Example of a CommentLabel

Figure 10 shows an actual example of a CommentLabel from experiment 2. The island name says "You should change the signboard at the gallery and you should arrange the detailed introduction in the gallery". It contains the comment "I do not attract attention. The sound of cars is unpleasant. I cannot observe it slowly. I wanted you to do a more detailed exhibition."



Figure 10: Example of a story that subjects made

## 6 DISCUSSION

### 6.1 The identification of photographs using position information

We now look at the frequency of use of the position lining up function in experiment 1. The use counts of the function were half in comparison with the number of islands. In addition, the evaluation average of item 1 of Table 2 was 4.4. We next discuss the effect of the function. We regard relations of the evaluation value of each subject for item 3 as item 1 of Table 2. We understood that the strong plus of 0.849 showed a correlation between the values of the two evaluations. Table 8 shows the discussion. Table 8 is a list of the correlation coefficients between each item in Table 1. Function practical use and the remark counts of the correlation coefficients are 0.939. This numerical value shows a very strong equilateral correlation. We believe that which promotes the activation of the argument is checked by the use of the function. The comparison of the working time of each group is - 0.713. This value shows a slightly strong negative correlation. We thought that shortening of the working time was possible if the use counts of the position

lining up function increases. In addition, the evaluation average was 4.2 (Table 2 (2)).

When we grouped the photographs which several people provided, we thought it possible to use digital photographs with position information for the identification of photographs.

Table 8: A list of the correlation coefficients between each item in Table 1

| Item | Work time | Number of comments | Position lining up | Enlargement |
|---|---|---|---|---|
| Work time | 1 | | | |
| Number of comments | -0.49326 | 1 | | |
| Position lining up | -0.71314 | 0.93984 | 1 | |
| Enlargement | -0.41957 | 0.93607 | 0.93048 | 1 |

## 6.2 Performing the grouping of photographs with several people

We next turn our attention to the number of remarks of each group and analyze the content of the remarks. Table 8 shows a correlation between the number of remarks and the number of times the enlargement function of photographs was used. As for this numerical value, a very strong equilateral correlation of 0.936 was identified. There was a remark "which group this photograph should be in". We discovered a correlation between the number of uses of the position lining up function and the number of remarks. The numerical value was 0.939, which we understood to be a very strong plus. We supposed that the argument became active when each function of the system indicated above was operated. Next, we looked at personal remarks between each group. A and D, which have small dispersion of remarks of subjects, have a short working time. In contrast, B, which has large dispersion of remarks of subjects, has a long working time. Figure 11 shows that we subtracted the personal remark counts from the average remark counts of each group. From Figure 9 and figure 11, there were many people who answered that it had "the merit of performing with several people" when the number of personal remarks was high. The correlation coefficient was 0.682.

The satisfaction of the subjects rises if they have lively arguments during the work of grouping PhotoLabels. We can suppose that working time is shortened because individuals speak equally.
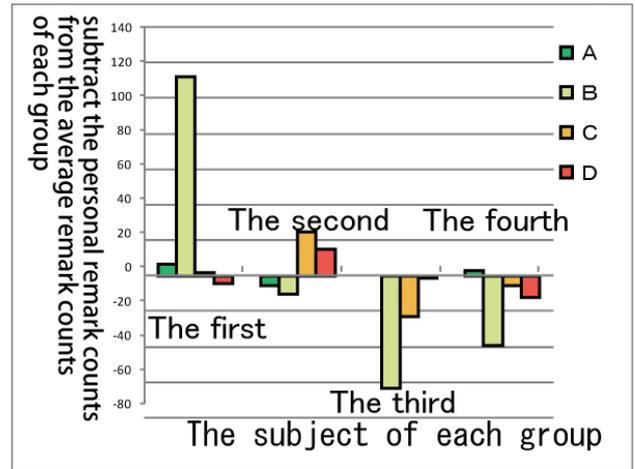


Figure 11: Subtract the personal remark counts from the average remark counts of each group

## 6.3 Making CommentLabels

We now describe the effect of the brainstorming that uses photographs with position information in article making in experiment 2. The evaluation value shows whether the CommentLabel function was effective (4.4). As for the CommentLabels, 1.5 (on average) were added for each island. Many ideas were proposed in CommentLabels. Some islands had no CommentLabels. In this case it is an island without a particularly strong impression. The number of cases of "an opinion and agreement" in each group increased by procedure 2, and opinions appeared more lively than in procedure 1. The ratios of the agreement remark counts among all remarks increased to 0.31 from 0.23. We supposed conversations arose between subjects by procedure 2.

## 6.4 Brainstorming using photographs with position information with several people

As for the number of islands, each group was the same as between procedure 1 and procedure 2. We thought that this is because procedure 1 involves choosing photographs at the time of making trip articles. However, there was only one example that was different in the photograph which was stored on an island. The total number of remarks for each group and item 1 of Table 6 correlate. The value was 0.6680 - about the same as experiment 1. In addition, we looked at the content of the remarks of each group by experiment 2 to examine which remarks were effective in procedure 2. The correlation coefficient value was highest for "opinions", at 0.5537. The other three items had almost constant values. Each value was 0.3129-0.3428. We can guess that merits for participants rise if numerous opinions appear at the time of making comment labels. Next, we compare procedure 1 in experiments 1 and 2. Members in each group unit could understand the meaning of all photographs, so the photograph lining up function was not used frequently.

# 7 CONCLUSION

We have developed GUNGEN-Photo. This system is an idea generation support system using digital photographs with position information to perform with the various people. We carried out 2 experiments on GUNGEN-Photo. The first experiment checked the effectiveness of identification by position information of digital photographs and grouping of photographs to perform with several people. The second experiment checked grouping of photographs to perform with several people (subjects took pictures by a group unit), effects of making CommentLabel and what kind of effect there was for idea generation. As a result, the following were understood.

(1)    We can support the identification of photographs by position information.
(2)    Many ideas were proposed in CommentLabels.
(3)    As for the number of island, each group was the same in procedures 1 and 2 together. The number of islands was 8.2.
(4)    The satisfaction of the subjects rises if they have lively arguments during the work of grouping PhotoLabels.
(5)    The ratio of "agreement remark" counts among all remarks increased to 0.31 from 0.23 in procedure 2 (CommentLabel adding step).

In the future, we would like to improve our software for the idea generation support system using digital photographs with position information by various people, and it is our aim to establish it as entertainment.

## REFERENCES

[1] A.F.Osborn,"Applied imagination: Principles and Procedures of Create Problem Solving (Third Revices Edition), *Charles Scribner's Son*, New York, NY,1963.

[2] J.Kawakita,"An Idea Development Method", Chuuko Shinsho, Chuuo Kouron-sha, 1968, Tokyo.

[3] H.Kawashima, K.Sugiyama,"Research and Development of Gropup Idea-Processing System Using a Mandala Figure", *The3rd Conference on the Support System for Knowledge Creation*,2006,pp.1-8.

[4] J. Munemori, I.Horikiri, Y.Nagasawa,"Groupwaere for New Idea Generation Support System (GUNGEN) and Its Application and Estimation to the Student Experiments of the Distributed and Cooperative KJ method, *Journal of Information Processing Society Japan*, Vol.35, No.1, 1994,pp.143-153.

[5] M, Ohashi, J, Ito, J, Munemori, M, Matsushita, M, Matsuda, Development and Application of Idea Generation Support System Using Table-top Interface (Collaboration /Recommendation,<Special Issue> Human-Centered Universal/Ubiquitous Network Services), ISSN2008.pp.105-115 (2008)

[6] J.Itou, K.Shiina. J.Munemori, "Development and Application of Electronic Memory Notebook to Japan and China", IPSJ SIG Technical Report 2008-GN68, pp.19-24(2008)

[7] K.Watanabe, K.Tsukada, M.Yasumura, "PhotoLoop: An annotation system using users' activities while watching slideshow", WISS2007, pp.93-96(2007).

[8] M, Ohashi, J, Ito, J, Munemori, "Proposal of an Creativity Support System Using Direct Manipulation on a Tabletop Interface", KICSS2008, pp.200-205(2008)

[9] K.Dietz and D.Leigh,"DiamondTouch:A Multi -User Touch Technology",*Proc.UIST.2001*,2001,pp.219-226.

# AR Chemistry : Building up Augmented Reality for Learning Chemical Experiment

Daiki Nishihama[†], Tatsushi Takeuchi[†], Tomoo Inoue[‡]and Ken-ichi Okada[†]

[†]Graduate School of Science and Technology Keio University, Japan
{nishihama, takeuchi, okada}@mos.ics.keio.ac.jp
[‡]Graduate School of Library, Information and Media Studies University of Tsukuba, Japan
inoue@slis.tsukuba.ac.jp

## Abstract

Learning is changing behavior and ideas by storing experiences and acquiring knowledge. We focus attention on Chemistry in various learning. This paper describes a learning assist system for chemical experiments "ARChemistry" which is used in augmented reality and tabletop interface for combining physical papers and digital content.

In this system, tabletop interface shows needed related information on tabletop interface. Augmented reality enables us to show chemical reactions by using a virtual model. The user's history is stored for the feedback of the experiment. ARChemistry's feedback function enable us realize mistakes that are made in the "Which procedure", "Who", and "What operation" aspects of the experiment.There are no methods to catch mistakes on traditional experiment.

Through experimental use, the results showed that an augmented reality interface system worked appropriately.

## 1 Introduction

Learning is to change thinking and actions by experience and storing knowledge, for example, remembering how to get to a new office, making a presentation or memorizing new words.

In this research, we focus on academic learning. There are many different types of learning. Collaborative learning is one type, in which, several people work together to achieve a common goal. Collaborative learning often takes place in chemistry courses during group learning experiments. We decided that this is one effective way of learning[5][6].

Many researchers have studied collaborative learning systems that use a tabletop interface or tangible device. Recently systems have been developed with large displays that allow two or more users to intuitively operate the device[1][3][2]. These systems allow users interact with physical objects and retrieve relevant digital information. This paper will propose "ARChemistry" should be used in introductory chemistry courses. The system assists the chemical experiment learning by using augmented reality (AR) and physical objects. The system has an experiment study space with the tabletop interface, and we apply that to chemical experiment. When students use the AR loupe for an experiment they will be able to use physical objects view 3D animations and retrieve any needed information like moving ion, chemical reactions. Augmented Reality offers a way to merge virtual graphics with the real world in real time[3]. Users will be able to see invisible objects with the system.

By Storing history information the system allows users to search for and find their mistakes in the "Which procedure", "Who", and "What operation" aspects of the experiment. With the conventional experiment it is impossible to refer to history information to find mistakes. It was impossible to refer the history information, but our system enables us to get the feedback. It is important for student to identify and isolate their mistakes. To confirm the utility of this system, an assessment experiment and user discussion were done.

## 2 Related Work

Learning is our ordinaries of life and it influences our operation afterward. Everyone has done it regardless of conscious or unconscious consideration. There are many different types of learning and an Academic learning is a conscious learning.

It falls into two general categories of learning that one is to learn from a textbook and a reference book or that another is to learn from the out-of-door activity and experiment.We focus on the experiment learning of the experience type[2].

### 2.1 Experiment Learning

We view the experiment learning as a collaborative learning that some facing people work jointly at the same time. The second learning style is class room learning that teacher stands in front of students. Learners often achieve high result with the collaborative learning more than with classroom study[5].

In this case, important points are evaluation to group and personal resposibility. Evaluation to group has all learners in group effects in aiming to achive high level. Working in them responsibility, learners have a great awareness and get a high result. Because two factors are direct motivations for learning. Based on these factors, we focus on the chemical experiment that many students do it at the same time and that chemical reaction is very impressive.

## 2.2 Problem of Experiment Learning

In this research, we make the virtual chemical experiment to assist chemical learning on the tabletop interface. In view of the following factors, we designed the system in order to do such the experiment as actual experiment.

- Cooperation with text study

- Experiment accident

- Experimenter's error in operation

- Experimental environment

First, it is difficult for people to always put all studied information in the head while experiment. But students need many chemical information in experiment to understand invisible reaction. For example, combining ion and electoron/other ions , features of cation, how to use instrument and so on. As a result, students often refer to the text or just do without right information.

Second, chemical accident and user's mistake will be caused even if we pay very attention to them.

Third, chemical experiment needs equipment mostly which is not enough at home.

## 3 ARChemistry

ARChemistry is a supporting system for chemical experiment study. The system applies 3DCG to virtual experiment space on tabletop. If the above-mentioned problem is based, it becomes an important point in the following four points.

- Visalization by AR

- Display related information in experiment

- The experimenter's identification

- Record of experiment procedure

At first, movement of ion and electron is ordinarily invisible but it becomes visible by AR. Visualization by AR adds the understanding more than an actual experiment[8][9]. Repeating is important in study. But the repeating in experiment needs many kinds of chemicals and needs experimental instruments which we don't have usually at home. Visualization allows us using chemicals and instruments in a repetitive manner. For example HCl or H2S in chemicals and the gas burner etc can be used. These reduce cost and risks. Additionally, this system leverages not only AR but also physical objects. It becomes possible not to lose the experiment feeling by operating physical objects like beakers.

Second, if we access a object, system displays some related information on tabletop. Mainly the information displays ionization tendency or contents of beakers and so on. This operation bridges the gap between text study and experiment study.

Third, identification and record allow users to search for and find mistakes. Correcting work is difficult in an actual experiment though some chance of study are ruined when the experiment is finished while not correcting a mistake. However, warning and feedback became possible by recording

each user's record. Warning is to inform them of the thing that causing to the accident when keeping experimenting or the experiment cannot be correctly ended while working on the tabletop interface. Feedback is to be able to confirm and isolate whether it made a mistake in "Who", "Which procedure", and "What operation" because the procedure and the user's correspondence are made the text sentences after working, it was not possible to do by an actual experiment.

We aspire achievement of construction that virtual experiment system has four necessary elements.
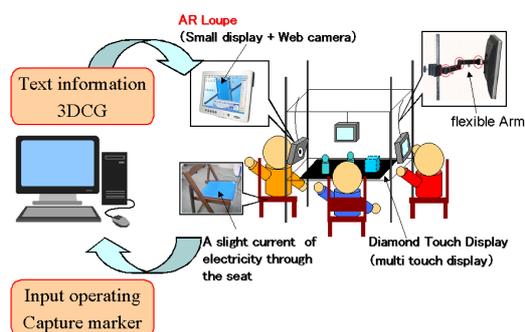


Figure 1: Hardware Construction

## 3.1 Implementation

This system keeps the horizontal and arranges a large-scale touch sensor display    DiamondTouch[4][7]    as shown in Figure 1. The electronic information is projected from the top of the rack, and the physical objects are on the display. We links the information and physical objects with tabletop interface. When user operates a beaker , the system displays beaker's contents information. In addition, we built up the AR loupe which has the Web camera behind a small display. We install the AR loupe by each user's left side. When AR loupe captured an AR markers, AR loupe will display some 3DCG about experiment which are images or some 3D animation. the system progresses order by order in the following three parts.

- operation physical object

- display AR

- operation with tabletop interface

It explains details and the manner of operation of these three parts in order.

### 3.1.1 Operation Physical Object

In this system, user can recognize the beakers as a lab ware. The experiment feeling is not ruined because we move the physical object on the touch sensor display, and the system can recognize it.

To recognize the physical object with the system, the copperplate and the copper wire with high conductance are put as shown in Figure 2.



Figure 2: Overview of Physical Object With Copperplate

A slight current of electricity from DiamondTouch passes the user's body when they operate a physical object, and it is read in the system[4][7]. DiamondTouch is the multi touch sensor display which can recognize the multi point and who use the system. Moreover DiamondTouch can distinguish the physical object according to the size of the sheet copper.

By using this function we can mix contents of the beakers virtually . As Figure 3, user grips the beaker firmly for eliminating a gap between hand and beakers. When we down the beaker on the tabletop , it was recognized. If the operation was correctly, included information would be shown on your space at Figure 4. And user select blending button when last two beakers have been recognized to the system. A marker includes text information and 3D animation, appears with sound effects. User can check the chemical reaction with AR loupe and text about experiment on tabletop.
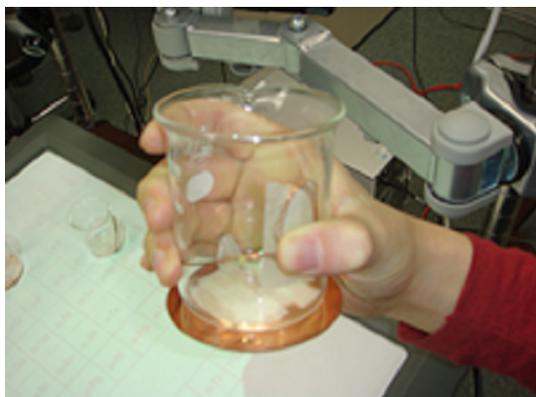


Figure 3: Gripping Physical Object
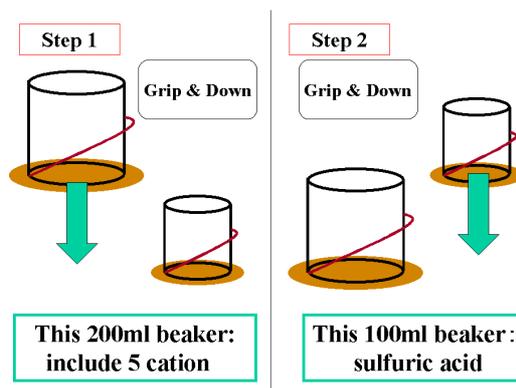


Figure 4: Physical Object Recognition

### 3.1.2 Display AR

We can use AR loupe for viewing virtual objects. As Figure 5, AR loupe is constructed by a 9 inch display and Web camera Qcam Pro for Notebooks . AR loupes are arranged for all three user's left side. Flexible arm fixes them strongly.
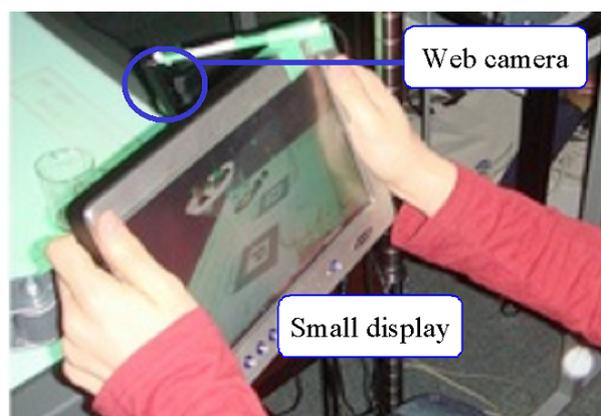


Figure 5: Operating AR Loupe

AR loupe distinguishes the presence of the AR marker through the image which Web camera captured. When the AR marker exists, loupe's display shows 3DCG on the marker after distinguish the shape. Moreover, because the AR marker has been linked not only 3DCG but also chemical information, textual information is shown on the tabletop and 3DCG is possible to display in loupe's display.

Because the AR loupe can be freely moved by using the arm as shown in Figure 5, it is possible to operate the system without closing both hands on the touch sensor display.

When users want to see 3DCG, they can pull AR loupe and peer into the screen of own small display. The display shows some animation. For example flame reaction or a precipitate by uniting ion and so on.

### 3.1.3 Operation with Tabletop Interface

We perform procedure chiefly by using the touch sensor display. Figure 6shows an actual operation screen.
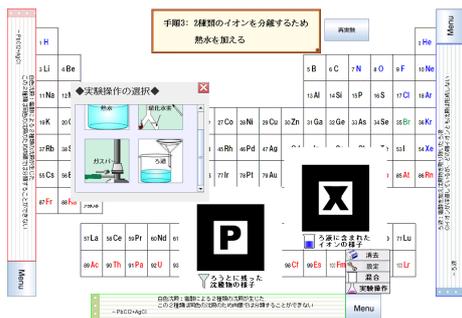
Figure 6: An Operation Screen on Tabletop Interface

It is possible to operate the system intuitively with touch sensor display. Additionally, We innovate a way to use menu list like ausual PC for handling so operating system is more easily. There are some panels to play the system on tabletop display.

1. an individual note pad panel
   When user operates the system, chemical information of operated object is shown to each user's front panel. The information will be shown in touching AR panel, in gripping a beaker, and after chemical reaction.

2. an indication panel
   This panel shows what you must do next. If you mistake the procedure, the system alerts that you are wrong with warning tone and message.

3. hiding button in menu list
   It hides the marker which user feels obstructive at operating it.

4. log button in menu list
   When this button is pushed, select panel appears. And on that panel, user decides whether they log or not, and what log file name is.

5. blending button in menu list
   As shown in Figure4, system memorize informations of two beakers. When the indication panel says that you must blend two beakers, this button is possible to use. At blending, AR panel shows 3D animation of chemical reaction process.

6. apparatus button in menu list
   In touching apparatus button, chemical apparatus panel appears as shown Figure 7. When user touches icon in this panel, system shows paired AR marker which corresponds to each operation such as lab wares and a reagent.Figure 8 seemed to select a gas burner icon, and to have looked into the marker with the AR loupe. When user touch again ,the gas burner reacts another flame color of the classified ion and user confirms whether it is a specific ion.

There are filtration, to pour heated water, the outbreak of the hydrogen sulfide as well as a gas burner.
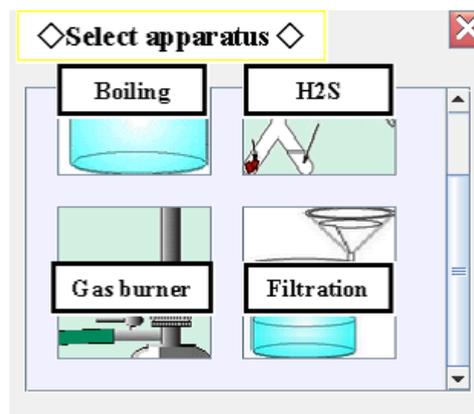


Figure 7: Overview of Apparatus Panel
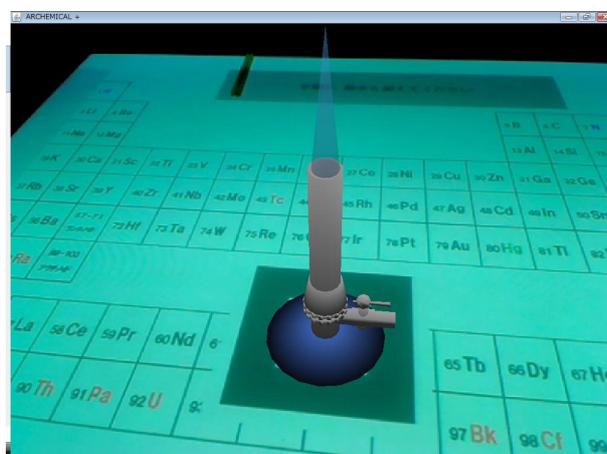


Figure 8: Apparatus　A Gas Burner for Flame Reaction

## 3.2　Record

When we do the experiment operation , our system records not only the operation but also a physical object and an operator. In a word, Users can confirm "Which procedure", "Who", and "What operation" aspects with the historical information whether they made a mistake.As a result, in actual experiments it was difficult to specify a mistake of procedure but our system allow users to discover a critical point in the experiment. And the system allow users to be able to prevent the accident before happens in real experiment becouse they are taught by warning message which shows that you must fail if you keep to do your operation.

## 4　Evaluation

The goal of this evaluation is to check the useful of our system.

## 4.1　Experimental Procedure

We did the experiment of cation analysis on our system. The cation analysis is chemical experiment that examinees isolate a certain cation from many kinds of cation in beaker.

Table 1: Summary of Questionnaire Response.

| No. | Question | Average |
|---|---|---|
| 1. | I found easy to manipulate table top | 8.0 |
| 2. | I found it intuitive to use beakers | 7.2 |
| 3. | I performed AR Loupe as intended | 7.4 |
| 4. | I imagined the actual chemical | 7.5 |
| 5. | I found it supporting experiment to use 3D or text infomation | 8.1 |
| 6. | I feel fun to use this system | 8.7 |
| 7. | I found Chemistry of interest | 6.5 |
| 8. | I found easy to understand chemical reaction | 8.1 |
| 9. | I found the system available for review and preparation | 7.7 |

Subjects has been done by the university student and post-graduates (14 men and 4 females) who did not make the chemistry a specialty. They filled in the questionnaire (ten stage evaluation and free description column). In the beginning after we explained the function of the research background, the outline, the manner of system's operation, user operated our system freely . Next they followed the instruction of the system and did "the cation analysis". After the experiment had been completed, users were required to fill the questionnaire.

## 4.2   Result

The Table shows the outcome of the experiment. When the summary is seen, the high reputation value is obtained for the question No.1, No.5, No.8 that we emphasized the question when we made this system. Responses say that it was easy because it was a very simple operation, I understood the appearance after it reacts and before it reacts well, and I obtained details of chemical reaction and the image of the experiment, it can be said that this system has utility. In addition, the question No.6 , No.9 was the high reputation values too. These comments are that experiment was so fun because we had been speaking each other while working and that it was possible to operate the system without getting tired. From this comment, the system provides understanding and interesting. Because efficiency goes up to study enjoying it, the evaluation backs up the utility of this system that these evaluations.

## 5   Conclusion

There are styles of the learning now for many divergences. Interaction methods with AR and a touch sensor display increased slowly in that. The system "ARChemistry" used AR and tabletop interface and put a focus for the chemical experiment learning. In this system, tabletop interface can show needed related chemical information on tabletop and how you push forward operation. Augmented reality allow us to show chemical reaction by using a virtual model. And the feedback of the experiment is enabled by storing user's history. The feedback is able to realize us the mistakes made "Which procedure", "Who", and "What operation" aspects.They are new method to combine text study and experiment one in chemical. Through evaluation experiment, the results showed that the system of an augmented reality interface worked appropriately and is interesting in learning and discussion each other.

## REFERENCES

[1] Diana Africano and Sara Berg, Kent Lindbergh and Peter Lundholm and Fredrik Nilbrink and Anna Persson : Designing tangible interfaces for children's collaboration, CHI '04 extended abstracts on Human factors in computing systems, pp.853-868 ,2004.

[2] H.Koike    Y.Sato and Y.Kobayashi: Integrating paper and digital information on EnhacedDesk: a method for real-time finger tracking on augmented desk system, ACM Trans. on Computer-Human Interaction, vol.8, No.4,pp.307-322,2001.

[3] Kikuo Asai and Tomotsugu Kondo1 : Molecular Structure Visualization System Using Augmented Reality Journal of Multimedia Aided Education Research 2008, Vol. 4, No. 2, 37     43

[4] Keigo Kitahara,Tomoo Inoue,Hiroshi Shigeno,Ken-ichi Okada: A Tabletop Interface for Supporting Collaborative Learning, IPSJ Journal, Vol.47, No.11, pp.3054-3062, 2006.

[5] Masaya Okada, Akimichi Yamada, Mizuki Yoshida, Hiroyuki Tarumi, Takanobu Kayugawa, Kazuyuki Moriya : Collaborative Environmental Learning with the DigitalEE II System Augmenting Real and Virtual Experiences IPSJ Journal,vol45-1,pp 229-243, 2004.

[6] Masaya Okada, Hiroyuki Tarumi, Tetsuhiko Yoshimura, Kazuyuki Moriya, Tetsuro Sakai : DigitalEE: A Support System for Collaborative Environmental Education Using Distributed Virtual Space,Systems and Computers in Japan, vol.33, No.8,pp 51-63, 2002.

[7] Dietz, P., Leigh, D., "DiamondTouch: A Multi-User Touch Technology", ACM Symposium on User Interface Software and Technology (UIST), pp.219-226, 2001.

[8] Moffat Mathews, Madan Challa, Cheng-Tse Chu, Gu Jian, Hartmut Seichter, Raphael Grasset : Evaluation of spatial abilities through tabletop AR CHINZ '07: Proceedings of the 7th ACM SIGCHI New Zealand chapter's international conference on Computer-human interaction: design centered HCI ,vol.254, pp 17-24, 2007.

[9] Fotis Liarokapis, Robert M. Newman:Design experiences of multimodal mixed reality interfaces Proceedings of the 25th annual ACM international conference on Design of communication,pp 34-41 , 2007.

# Visualizing Liveliness of Discussions and Reply Relationships
# on Online Discussion Boards

Hidekazu Shiozawa

College of Engineering, Tamagawa University, Japan
shiozawa@eng.tamagawa.ac.jp

*Abstract* - This paper proposes a technique to interactively visualize the liveliness of discussion and the reply relationships of messages on online discussion boards, and describes the prototype software that visualizes liveliness of any topic thread on 2channel (Ni Channeru) [1], the Japan's largest online discussion site. The visualization is based on a line graph of the number of postings versus time. To show reply relationships on the same chart, this visualization utilizes the technique of Thread Arcs [14] that shows reply relationships as the semicircle arcs connecting between message plots. Furthermore, for users to know the length of each message, the visualization can show plot a message as a pair of small bar graphs on the same line graph.

*Keywords*: liveliness of discussion, visualization of discussion structure, online discussion boards, 2channel (2ch.net), Thread Arcs

## 1   INTRODUCTION

An enormous amount of information is exchanged every day in online discussion boards or Web forums in the Internet. Though such sites are mixed bags of various information, lively parts of discussions in which many people participate would likely include some valuable, interesting or fun information. However, conventional discussion boards provide only user interfaces that users have to read messages one by one from the beginning of each thread, even if they are looking for only lively parts of the discussions.

Therefore, this paper proposes the technique to interactively visualize the liveliness of discussion and the reply relationships of messages on online discussion boards, then also describes the prototype software that visualizes the liveliness of any topic thread on 2channel (Ni Channeru) [1], the Japan's and probably world's largest online discussion site. Although the "liveliness" of discussion can be considered to relate various factors of the discussion, this paper roughly defines it as "the number of posted messages per unit time", and the number of replies per message and other statistical data are used as secondary information.

The 2channel consists of many Web discussion boards for different themes, and each board allows a user to make a new "thread" for every new topic. In each thread, posted messages are displayed sequentially from 1 up to 1000 in message number. In each board, displaying order of its threads is determined by floating the thread with a new posting to the top, while the user may use the "sage" (down) command to keep the thread from floating.

In addition, 2channel features its anonymity[1] that users are not required any registration or contact information. Actually, most of the postings are contributed by anonymous. However, recently to reduce the negative effects of the anonymity, many boards append a string called "ID", which seems a random string, to each message for roughly identifying the user. The ID string is calculated with a one-way hash function from both a posting IP address and a posting date, so it is not changed while the user post any message from the same address on the same day.

## 2   RELATED WORK

### 2.1   Analysis of Online Discussion Boards

As for the liveliness of the online discussion boards, Matsumura et al. [2], [3] collected and analyzed log data from all discussion boards in 2channel to propose the quantification of characteristics of each discussion thread according to the following eight indices.

**Contents (C)**  The average size of a message in bytes (except the following AA and V).
**Activity (A)**  The average number of messages per thread.
**Interactions (I)**  The average number of replies per message.
**Speed (S)**  The average number of messages per day.
**Vocabulary (V)**  The byte ratio of "2channel words".
**Ascii Art (AA)**  The byte ratio of text arts.
**Nameless (N)**  The ratio of the messages posted by anonymous ("nanashi-san").
**ABON**  The ratio of the messages deleted[2] by administrators.

They calculated these indices by category, which discussion boards belong to, and show the rough trends of discussions in each category by analyzing the relation of these indices with the method of covariance structure analysis. Then, they proposed the statistical model to show the overall dynamism of 2channel with the cause-effect relationships of the three trends: the discussion divergence trend (A and S are large), the discussion deepening trend (C and I are large), and the stylized expression trend (V and AA are large).

In addition, Matsumura et al. [5] also performed statistical analysis for the dynamism of Yahoo! Japan Message Board [4], which requires user's registration unlike 2channel. They used the weighted directed graph representing relation of users to analyze reply relationships and their influences in a discussion. Then, they focused on the role of users and

---

[1]Indeed, it is not so anonymous, because the servers maintain the log data of all connections.
[2]"Abon" is the system message indicating deletion of the message.

classified discussions in Yahoo! Japan Message Boards into the three types: the leader-driven communication, the leader-follower-collaborative communication, the follower-driven communication.

These researches investigated users' collective behavior in online discussion from macro perspective by quantifying a discussion with several index values. Therefore, the techniques can be significant in the fields of sociology, mass psychology and so on; however, they are not directly useful for general users reading online discussion boards.

## 2.2 Time-Series Analysis of Discussion

From users' point of view, discussion boards with the more postings are considered the more worth reading, because usually such boards hold lively and active discussions and abundant fresh information. As for 2channel, the degree of discussion liveliness is often called "ikioi" (momentum) or "nobi" (growth) [3]. These words are considered to roughly mean the number of postings per unit time ("growth" also means the total number of postings).

For users to find such lively and interesting discussions, some software [6], [7] and Web sites [8], [9] for reading online discussion boards display the value called "ikioi" (momentum) or the posting speed per hour for every discussion thread, which is the average number of postings per time. Moreover, some show ranking lists of the values.

Furthermore, there is the software called V2C [10], which can visualize the time variation of the value on a chart. V2C draws a thread in a line graph using x-axis for time and y-axis for message number (i.e. the number of postings). Therefore, the slope of the graph represents the number of postings per time or the momentum.

Matsumura et al. [11] extended the research previously mentioned [2] and tried the time-series analysis of the liveliness of a discussion in 2channel which the following new six indices: the posting interval rate (the average interval between postings), the interaction rate (the average numbers of replies per message), the anonymous rate (the rate of anonymous postings), the 2channel word rate (the rate of jargons per message bytes), the ASCII arts rate (the ratio of characters often used in text arts), and the new information rate (the rate of new nouns per message).

## 2.3 Visualization of Discussion Structure

Visualizations of discussion structures for character-based online communication have been developed and used conventionally. Most of the software for email and Netnews (Usenet) enable users to gather a related discussion and to view relationships of messages in a tree structure by using the messages' headers such as "Subject" and "References". This method visualizes a lively discussion in Netnews as a complicated tree structure because of its message format like email.

In the CSCW field, many discussion systems [13] with visualization interfaces have been proposed traditionally. Most

---

[3]Moreover, the state of extremely lively discussion is sometimes called "matsuri" (festival), where users join the discussion and post many messages in aim of liveliness itself.
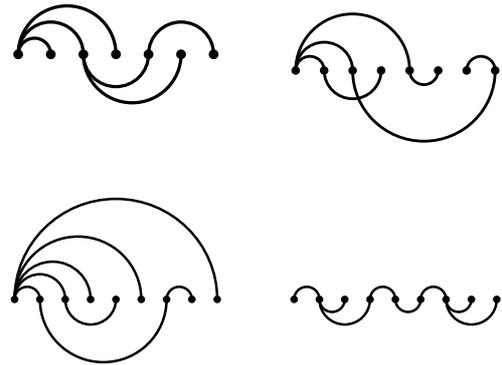
Figure 1: Thread Arcs (B. Kerr, IBM Research)

of such systems show the discussion flows with branching and merging as directed graph structures on the two-dimensional plane. For example, Matsumura et al. [12] proposed the analysis method and the visualization of the discussion structure on their online discussion board as a kind of directed graph.

On the other hand, this paper refers to the research of Thread Arcs [14], visualization for reply relationships of email in a personal mailbox. As Figure 1 shows, the Thread Arcs visualization can show reference relations among messages placed on the straight line, by drawing semicircle arcs connecting replying messages and replied ones. As a result, the compact visualization that does not need two-dimensional layouts is realized.

## 3 INFORMATION TO BE VISUALIZED

As suggested by the words "ikioi" (momentum) and "nobi" (growth), it is considered that users' feeling about the liveliness of online discussion is greatly affected by the passage of time. Therefore, this research uses the visualization based on a statistical graph with time axis to display the following information.

**The cumulative number of postings** Visualizing this is the most basic view for liveliness of discussion over time. The line graph like V2C is used. It relates Matsumura's A and S indices.

**Postings by the same ID (user)** The messages posted by the same ID as currently focused by the user are highlighted in the view.

**The cumultive number of unique IDs** If this is small, it means that few users perform many postings.

**Postings with "sage" command** Posting messages with "sage" (down) command keeps the thread from floating to the top of the board.

**Size of every message** The visualization can show the size of each message and that except symbols and spaces.

**Reply relationships between messages** Every replies in the thread is visualized. It relates Matsumura's I index.

**The cumultive number of replies** It relates also Matsumura's I index.
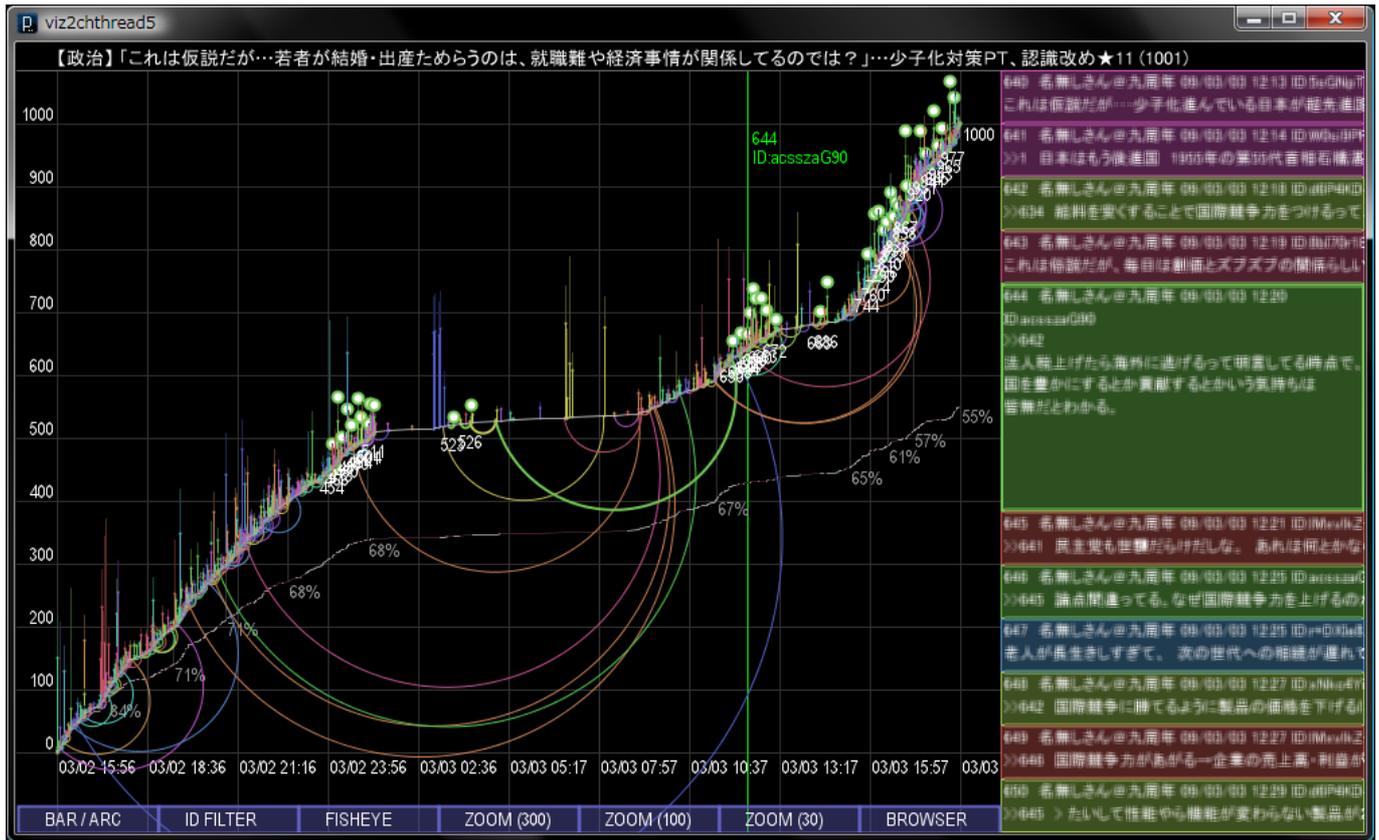
Figure 2: The visualization overview (the left area is for visualization, the right area is for messaeg texts)

For now, there is no metrics corresponding to the number of 2channel words (V), the rate of anonymous postings (N) and the number of deletions by administrators (ABON) in Matsumura's indices [2], [11]. The author considers that these are not so important for interactive visualization for a single thread, because most of the new jargons are used frequently only in short periods and the numbers of signed postings and deletions of messages are small in a thread.

On the other hand, this research uses information about ID and "sage" (down), though these are not included in Matsumura's indices. These are considered useful to characterize each thread, because the number of IDs can be a useful index for the number of unique posters and "sage" postings are seen commonly in threads that participating users are little changed.

In addition, this research does not count the replies to the first message, which presents the topic of entire thread, as an effective reply.

## 4 VISUALIZATION TECHNIQUE

This section describes the visualization technique in detail. The prototype software has been developed with Processing [15], the programming language suitable for dynamic graphics. When a user drag-and-drops the log file or the URL of the thread, which is saved by a 2channel browser like Open Jane [6] or displayed on a Web browser like Firefox, onto the window of the visualization, the screen as shown Figure 2 is displayed [16], [17].

### 4.1 Line Graph for Discussion Liveliness

This visualization technique is based on a line graph of the cumulative number of postings versus time. By this, user can read the total number of postings until some point and the process of making the discussion lively by the slope. If the slope of the graph is large at some point, the discussion at the time is very lively, and if small, it is not so lively.

Each posted message is plotted with a different color (hue) calculated from its ID on the graph. When a user selects a message to investigate its relationships, the messages posted by the same ID are highlighted on the chart with particularly large plots (Figure 3 and 4). These features enable users to find posting tendencies of other participants.

### 4.2 Line Graph for the Number of Users

The software visualizes the cumulative number of unique IDs also as a line graph, which shares the vertical axis with the number of postings, and shows the percentage of unique IDs for all postings in digits every hundred messages. If the percentage is small, some regular users are continuously posting in the thread; on the other hand, if large, many new posters are participating constantly.

Furthermore, the parts of the graph that messages are posted by "sage" (down) commands are drawn in thinner color. Usually threads sinking in this manner are hard to be found by newcomers on the discussion boards.
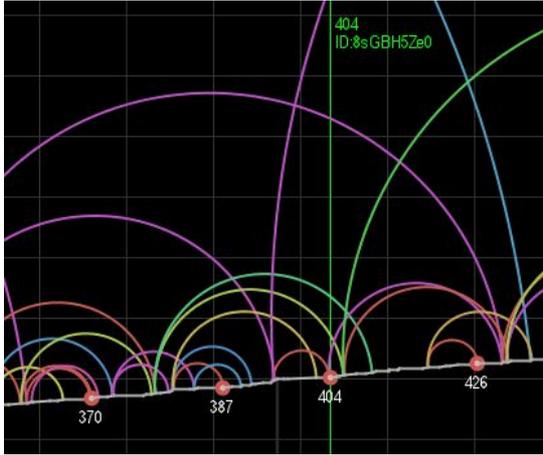
Figure 3: Arcs for reply relationships and coloring for IDs



Figure 4: Small bar graphs for the amount of information

## 4.3 Small Bar Graphs for Message Size

The software can visualize the approximate amount of information contained in each message with the two small bar graphs at a position of each plot on the line graph of the number of postings (Figure 4). These two graphs represent the length of each message in bytes and the length except symbols and spaces in bytes respectively. The latter's bar is drawn more thinly than the former's.

If the lengths of the pair of bars are largely different on a plot, the corresponding message contains many symbol characters. Therefore, this feature enables a user to estimate the effective size of every message except text arts, formatting spaces, and so on. Although Matsumura removed text arts with the dictionary for calculating the effective size of every message, this interactive visualization supports users to know the amount of information visually from the bar graphs.

## 4.4 Arcs for Reference Relationships

The feature of this interactive visualization is to display reply relationships of messages in semicircle arcs on the same line graph of the number of postings by using the technique of Thread Arcs (Figure 3). As a result, a user can investigate the tendency of replies in the entire thread, find the messages that have especially many replies and read the sequences where the discussion becomes much lively with many replies.

Applying the technique of Thread Arcs to an existing monotone increasing graph enables to visualize relationships of data as additional information on the same graph, because the technique visualizes connections between two plots on one dimension.

The arcs are shown on the upper side of the line graph when the small bar graphs are not drawn, while shown on the other side when the bar graphs are drawn to avoid visual overlapping. Their colors are same as the plots of the corresponding reply messages.

In addition, the cumulative number of replies is also visualized as a line graph like the number of postings and the number of unique IDs. This number often increases when some users in the thread have opposite views, for their hot debate brings many replies.
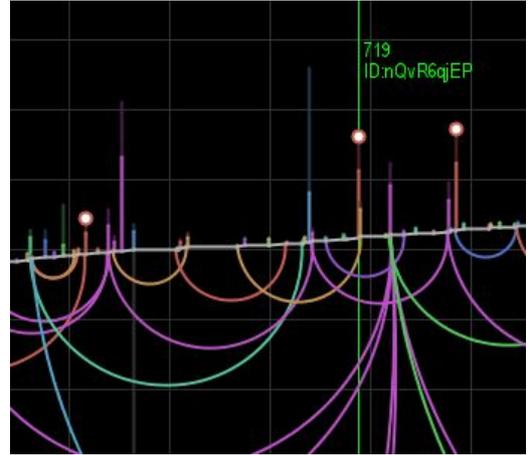
## 4.5 Partial Zooming and Fisheye View

The time axis of the graph ranges from the thread creation to the latest post; the vertical axis ranges from 1 to 1000 of message number. Plotting up to a thousand messages in a chart can make the visualization too dens and occluded, so the zooming feature is provided to show the range between the last ten and the next ten messages around the user's selected one.

However, some reply relationships are connecting temporally separated messages in a thread, so only clipping and zooming a part of a chart may be insufficient for a user to know relationships in detail. Therefore, the visualization provides a typical fisheye view with the technique of the Graphical Fisheye View [18] (Figure 5). It is also possible to apply with the simple zooming at the same time.

When you select a message in the mode of fisheye view, the $x$-coordinate is recalculated interactively by the formula,

$$x_{feye} = \frac{(d + 1)\, x_{max}\, x}{dx + |x_{max}|},$$

where $x$ is the $x$-coordinate before the transformation when the $x$-coordinate of the focus is zero; $x_{feye}$ is the $x$-coordinate after the transformation; $x_{max}$ is the $x$-coordinate of the display limit in zooming direction (depending on the sign of $x$); and the $d$ is the zooming parameter ($d = 8$ in the figure).

By this, a user can view the overview of the entire thread and the detail around the selected message and know roughly the destination point of arcs representing reference relationships.

## 5 APPLICATION AND DISCUSSION

This technique enables users to visually understand tendencies of individual threads in temporal view, though Matsumura et al. clarified tendencies of online discussions in a macro perspective with the several quantified indices.

This section discusses the several characteristic tendencies of threads found by this visualization in three perspectives: thread on event vs. thread on theme, thread with impressions vs. thread with debate and thread for open discussion
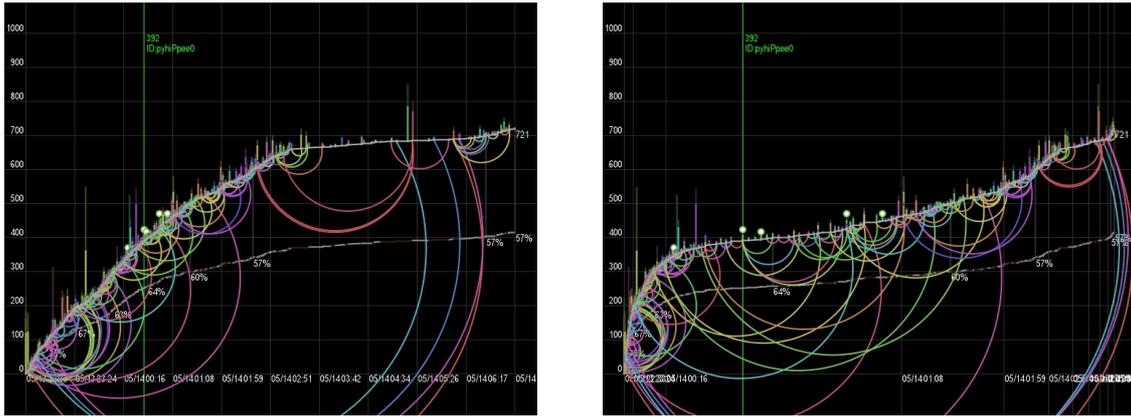
Figure 5: Nonlinear zooming with fisheye view, on the left is before applying and the right is after

vs. thread for members. These were found by applying this visualization to various types of threads in 2channel.

## 5.1 Thread on Event vs. Thread on Theme

By observation of the time variations of liveliness on many different threads from creation to end, some threads become most lively at their early stage from creation then are gradually decaying over time, and others become lively at random times regardless of creation and end.

Such phenomenon is considered to relate the purpose of each thread, or the reason of its creation. The former type is often seen when a kind of event triggers a user to create a thread to discuss on it; the latter type is often seen when the thread is continuously   maintained by some theme, such as a local area and some sports team.

For example, the right of Figure 6 is the thread of the breaking news about the national debt balance (for a day); the left is the different thread on the topic of the Kingdom of Thailand (for three years). The latter has been maintained in a variety of topics related to Thailand, but the tremendous surge of the coup d'etat is shown in comparison with before or after it.

## 5.2 Threads with Impressions vs. Threads with Debates

In many threads, some are lively on some level in the number of postings, even though there are very few replies in the thread. In contrast, some are not so lively and the increase of postings is slow, but discussion is performed actively with long messages and many replies.

One of the factors causing these differences is considered whether the topic of the thread is easy to debate or not. On some kinds of news, there is not much room for debates among users than reported in the first message, so the discussions do not become so interactive, even if the people post many messages of impressions such as surprise and repulsion.

Moreover, when the first message reports so shocking news, occasionally a number of messages without any replying comments are posted much rapidly, for example, simple messages with surprises, praises, opinions and agreements to the first message. It depends on readers to deem such communication as lively and important.

On the other hand, social problems, historical issues, and topics involving users often expand discussions to various directions and accelerate lively communications among users. It enlarges the number of replies in the thread.

For example, Figure 7 shows two visualizations for threads both in the newsflash board; the left is on the news of the prime minister's remarks and the light is on the news of exhibition of the Dead Sea document. The number of postings is not much difference, while the numbers of replies are in the contrast numbers, the former for 15% par the number of postings, the latter for 48%.

## 5.3 Threads for Open Discussion vs. Threads for Regular Members

It can be said that the threads with high percentage of unique IDs are open for discussion and various users are continuously participating and immediately leaving. On the contrary, the threads with low percentage of unique IDs are for some regular members and the relatively small number of participants continues to post messages.

The threads for regular members often continue to progress with "sage" (down) postings, because such threads are not so interesting for many other people. Although there is the small number of new participants in such thread, sometimes the discussion becomes very lively as an explosion. It is often triggered by news of the event related to the topic of the thread, such as results of sports matches, and indicates that the number of postings per member is increasing.

For example, the left of Figure 8 shows a thread on certain local news of a city, and the right is the thread for fans of the specific professional baseball team. The former indicates over 80% of the number of unique IDs, while the latter is 31% with keeping from floating with "sage" progress.

## 6 SUMMARY

In this paper, the author has proposed the technique to interactively visualize the liveliness of discussion and the reply relationships of messages on online discussion boards. This visualization consists of a line graph of the cumulative number of postings versus time to visualize the liveliness of dis-
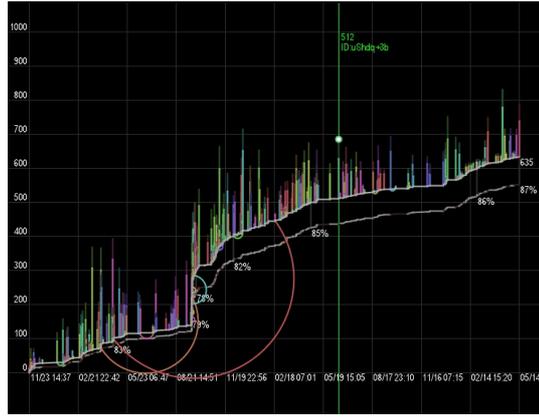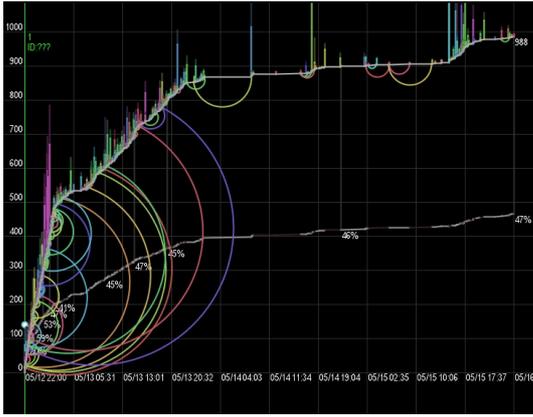
Figure 6: Visualizations of the threads on an event (left) and on a theme (right)
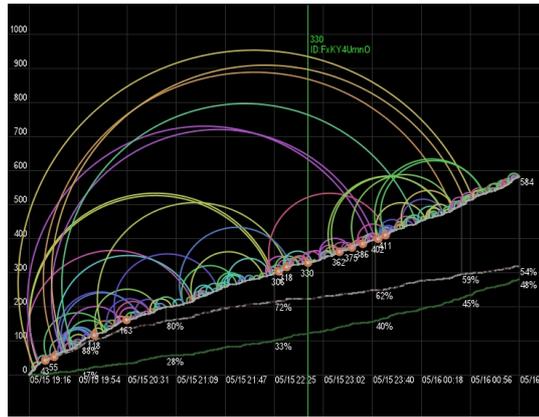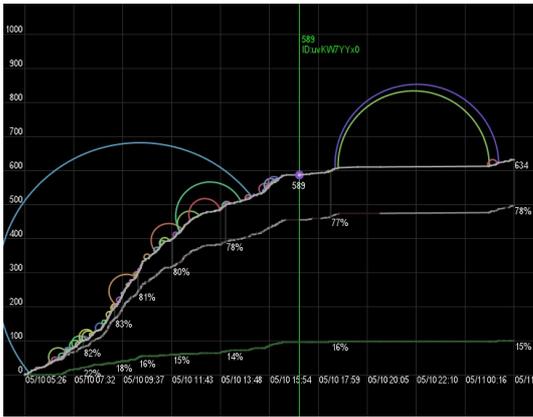


Figure 7: Visualizations of the threads with impressions (left) and with debates (right)
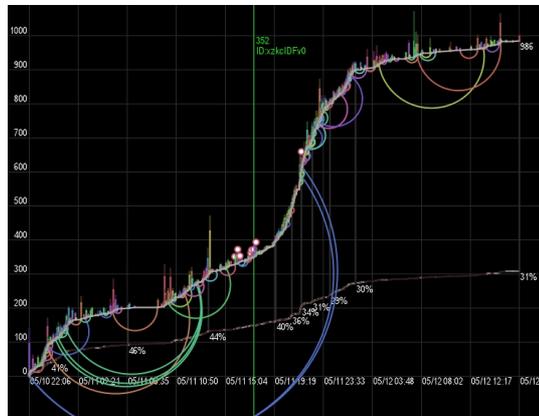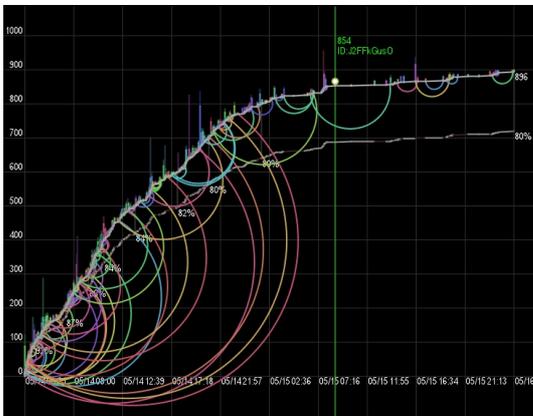


Figure 8: Visualizations of the threads for open discussion (left) and for regular members (right)
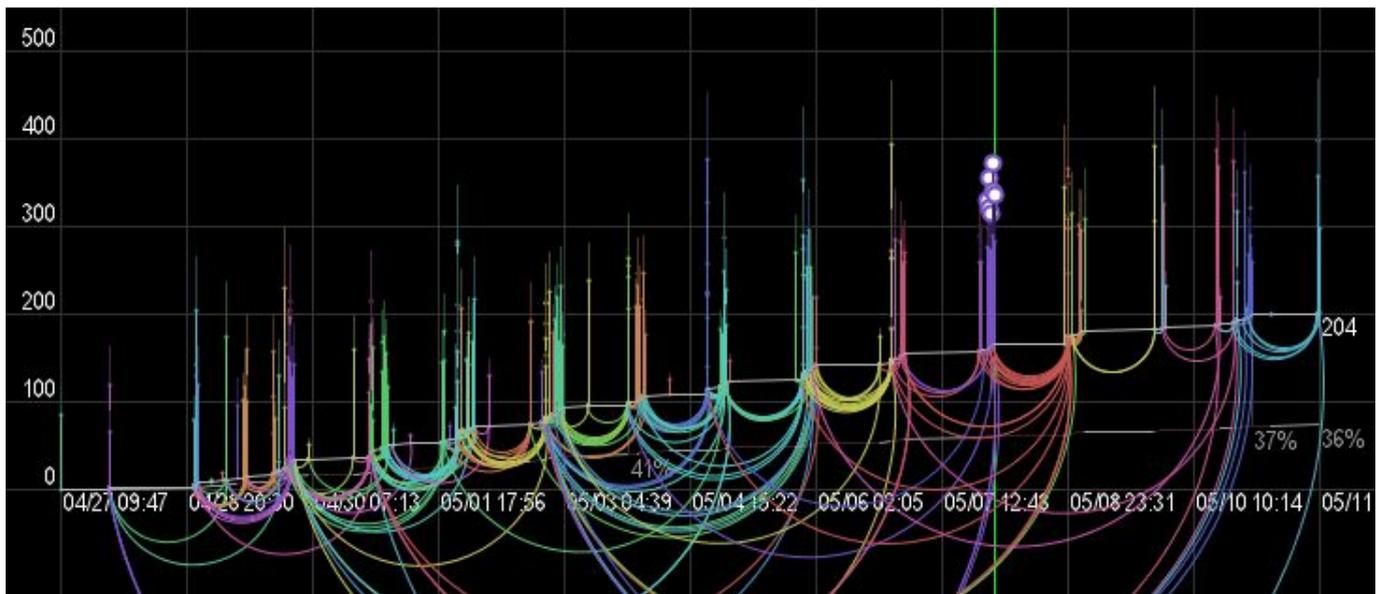
Figure 9: An example of very high correlation between discussion liveliness and time of day

cussion, and semicircle arcs connecting replying pairs of messages on the same graph.

In other words, this research extends the technique of Thread Arcs to visualize relationships of two points on a monotone increasing polygonal line or curve, though originally it can visualize only on a straight line. Therefore, this visualization is considered applicable to other existing line graphs to visualize additional information among data.

The author applied the method to the real threads in 2channel, and then the various characteristics in deferent threads were found. Thus, the author discussed them in the three perspectives: thread on event vs. thread on theme, thread with impressions vs. thread with debate and thread for open discussion vs. thread for regular members.

Some topics are considered for future work, such as experimental evaluation, visualization with effect of time of day, and feedback to the technique to statistical analysis. Three-dimensional visualization such as the Natto View may be useful to provide visual interaction on a chart [19].

Regarding the effect of time of day, many threads are affected by the human life rhythm, so it is known that the number of the postings is decreasing from a midnight to the next morning [16]. Furthermore, there are threads that have very high correlation between their liveliness and time; Figure 9 shows such thread that many messages are posted at the same time everyday.

## ACHKOWEAGEMENT

## REFERENCES

[1] 2channel (Ni Channeru), http://www.2ch.net

[2] N. Matsumura, A. Miura, Y. Shibanai, Y Ohsawa, M. Ishizuka: The Dynamism of 2channel, Transactions of Information Processing Society of Japan, Vol.45, No.3, pp. 1053–1061, 2004. (in Japanese)

[3] N. Matsumura, A. Miura, Y. Shibanai, Y. Ohsawa, T. Nishida: The Dynamism of 2channel, AI & Society, Vol. 19, No. 1, pp. 84–92, Springer Verlag, 2005.

[4] Yahoo! Japan Message Boards, http://messages.yahoo.co.jp

[5] N. Matsumura: The Dynamism of Yahoo! Japan Message Boards, Proceedings of The 19th Annual Conference of The Japanese Society for Artificial Intelligence , 2C1-02, 2005. (in Japanese)

[6] Open Jane Project Top Page, http://sourceforge.jp/projects/jane/, 2002-.

[7] noby: 2ch Hot Thread Search, http://www.forest.impress.co.jp/article/2006/03/01/ 2ch_hotthread.html, 2006.

[8] The Thread Ranking, http://www.bbsnews.jp

[9] R. Nakajima: 2channel News Navigator (2NN), http://www.2nn.jp, 2004-.

[10] n|a: V2C, The 2ch Browser using Java+Swing, http://v2c.s50.xrea.com/, 2004-.

[11] N. Matsumura, Y. Shibanai, H. Nishimura, A. Miura, T. Nishida: Time-series Analysis for the Communications on 2channel, Online Proceedings of The Fourth MY-COM, The Japanese Society for Artificial Intelligence, 2003. (in Japanese)

[12] N. Matsumura, Y. Kato, Y. Ohsawa, M. Ishizuka: Visualization of Discussion Structure for Discovering and Understanding Discussion Points, Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, Vol. 15, No. 5, pp. 554–564, 2003. (in Japanese)

[13] P. Kirschner, S. Buckingham-Shum, C. Carr (ed.): Visualizing Argumentation – Software Tools for Collaborative and Educational Sense-Making, Springer, 2003.

[14] B. Kerr: THREAD ARCS: An Email Thread Visualization, Proceedings of The Ninth IEEE Symposium on Information Visualization, pp. 211-218, 2003.

[15] C. Reas, B. Fry: Processing: A Programming Handbook for Visual Designers and Artists, MIT Press, 2007 (http://processing.org).

[16] H. Shiozazwa, H. Nakayama: Visualizing Liveliness of Discussion and Reference Relations on Online Discussion Boards, Proceedings of DICOMO 2008, Information Processing Society of Japan, pp.1308-1315, 2008. (in Japanese)

[17] H. Shiozazwa: Visualization of Liveliness of Discussion and Reference Relationships on Online Discussion Board, Proceedings of Interaction 2009, Information Processing Society of Japan, 2009. (in Japanese)

[18] M. Sarkar, M. H. Brown: Graphical Fisheye Views of Graphs, Communications of the ACM, Vol. 37, No.12, pp.73–84, 1994.

[19] H. Shiozawa, H. Nishiyama, Y. Matsushita: The Natto View : An Architecture for Interactive Information Visualization, Transactions of Information Processing Society of Japan, Vol. 38, No.11, pp. 2331–2342, 1997. (in Japanese)

# Analysis of Relationships between Atmosphere and Smiley during Plain Text Chat

Junko ITOU[†], Ayumu KIDO[†], and Jun MUNEMORI[†]

[†]Faculty of Systems Engineering, Wakayama University, Japan
{itou, s095016, munemori}@sys.wakayama-u.ac.jp

*Abstract* - In this research, we analyze the relationships among smiley, user's mental states, nonverbal expressions, and atmospheres during text chat aiming to apply to embodied character chat system. Smiley plays an important role in text communication to express user's emotions and dialogue atmosphere. We performed an experiment to get messages with smiley and investigated the relationships.

*Keywords*: Chat, Smiley, Emotion, Atmosphere, Dialogue, Embodied Character

## 1 Introduction

Online communication is widely spreading and tools of online communication become diverse, for example e-mail, chat system, remote meeting system, distance learning, and so on. There are also varieties of proposed tools on chat system from conventional text-based one to graphical one that agents in place of users talks in virtual 3D space.

By using embodied character in chat system, the users can obtain messages by watching embodied character's actions as well as by reading plain texts. A character in the chat systems plays a role as an agent of a user not only to express the user's emotional states or intentions which cannot be displayed by a chat message, but also to make the chat alive. As the result, it becomes clear what meaning the user implies for the chat messages.

In order to employ embodied characters for chatting users, we need to control their nonverbal expressions which include gestures, eye-gazes, noddings, and facial expressions, and so on. And also we should recognize the meanings of smiley which are used mainly to express the user's emotional state by text, furthermore we should reflect the meanings in nonverbal expressions.

Previous work on controlling nonverbal expressions of embodied characters mainly discusses the consistency of nonverbal expressions with the speech utterances or the goal of conversation for each agent [1][2]. However, when we consider dialogues between a pair of embodied characters, we need to consider interdependences between nonverbal expressions displayed by those two characters. As explained in more detail in the next section, it is reported in the field of social psychology that nonverbal expressions given by humans during their talks are not independent with each other[3]-[6].

In the remainder of this article, we focus on the relationships between atmosphere and smiley. Then we will discuss the chat atmosphere in terms of relationships among nonverbal expressions.

This paper is organized as follows: in section 2, we will describe the knowledge about the interdependences between nonverbal expressions in human communication reported in the previous work on social psychology. In section 3, we will define the "atomosphere" and the amount of nonverbal expressions in text chat. We will show the experimental result on the way of smiley and the association between smiley and atmosphere in chat in section 4. Finally, we discuss some conclusions and future work in section 5.

## 2 Nonverbal expressions observed in human conversation

In order to employ embodied characters for chatting users, we need to control their nonverbal expressions. As far as characters have their own faces and bodies in order to be "embodied", their users read various meanings in the nonverbal expressions displayed by the faces and the bodies of the characters even if the characters are not actually designed to send nonverbal expressions to their users but only to speak. Thus, for any embodied characters, we need to control their nonverbal expressions properly so that they convey appropriate meaning to the users. For example, it would be strange if a character does not smile at all although the user tells a funny story. As another example, when one character talks or smiles to the partner character, if the partner character freeze with no response to the action, it looks also strange.

It has been investigated in social psychology what features are found in nonverbal expressions given by humans during their conversation. Through those investigations, it is known that nonverbal expressions given by humans in real conversation have some interdependences and synchronicity.

For example, it was reported that the test subjects maintained eye contact with their partners in lively animated conversation, whereas they avoided eye contact when they are not interested in talking with their partners [3][4]. In the experiments by Matarazzo, noddings by the listeners in conversation encouraged utterances of the speakers, and as the result, animated conversation between the speakers and the listeners is realized [5]. In another experiments by Dimberg, fa-

Table 1: Interdependences between nonverbal expressions by different persons during conversations

| | | person A | | | |
|---|---|---|---|---|---|
| | | gaze | smile | nodding | speech |
| person B | gaze | * | | | * |
| | smile | | * | | * |
| | nodding | | | | * |
| | speech | * | * | * | * |

cial expressions of the test subjects were affected by those of their partners[6]. The subjects smiled when their partners gave smiles to them, whereas they gave expressions of tension when their partners had angry faces. Table.1 illustrates these positive correlations by *.

These results implies positive correlations or synchronicity between the nonverbal expressions given by the conversation partners for eye gazes, noddings and facial expressions. On the other hand, these interdependences are general tendencies in the usual conversation situation and can change by the second according with the conversation. Therefore, we can say that the strength of these interdependences are different by the atmosphere of the scene of talks.

However, it is difficult to convey these nonverbal expressions and relationships in plain text communication. To express their emotions and intentions, people substitute smiley for nonverbal expressions in the text communication. We focus on this smiley in chat messages.

## 3 Chat text and nonverbal expression

### 3.1 Definition of atmosphere function

We consider chat messages as utterances in face-to-face communication and define variables as follow:

- Utterance: the number of words in one message

- Thinking: the time for reply

- Emotion: the kind and the intensity of smiley

We define "atmohsphere" as a level of "chat liveliness" by the amount of above variables. When the two persons' "Utterance" is high, "Thinking"is short, and "Emotion" appears frequently, we regard the atmosphere level is high, that is "the chat is lively". On the other hand, when "Utterance" is low or "Thinking" is long or "Emotion" rarely appears, the level of atmosphere is low. Additionally, we consider the chat atmosphere level is low if two persons' total variables don't have similar value, because of the incoherent state.

In order to estimate the state of atmosphere, we define some variables and functions. We denote each value of Utterance, Thinking and Emotion, which the chat message of person $P$ at the time $t$ by $e_P^u(t), e_P^h(t), e_P^s(t)$, and we formulate atmosphere contribution function $A_P^u(t), A_P^h(t), A_P^s(t)$ with these variavles as eq.(1) $\sim$ eq.(3).

$$A_P^u(t) = \begin{cases} 0 & (e_P^u(t) < m_P^u) \\ 1 & (e_P^u(t) \geq m_P^u) \end{cases} \quad (1)$$

$$A_P^h(t) = \begin{cases} -1 & (e_P^h(t) \geq 60sec) \\ 0 & (e_P^h(t) \geq m_P^h) \\ 1 & (e_P^h(t) < m_P^h) \end{cases} \quad (2)$$

$$A_P^s(t) = \begin{cases} 0 & (e_P^s(t) < 1) \\ 1 & (e_P^s(t) \geq 1) \end{cases} \quad (3)$$

eq.(1) means the atmosphere contribution function for Utterance $A_P^u$ has the value of 1 when $e_P^u(t)$ that is the number of words in one message of person $P$ is bigger than $m_P^u$ that is the average of the number of words in one message of person $P$. In the same way, eq.(1) means the atmosphere contribution function for Thinking $A_P^h$ has the valeu of 1 when $e_P^u(t)$ that is the time for reply from person $P$ is shorter than $m_P^h$ that is the average time for the reply in one message of person $P$. The atmosphere contribution function for Emotion $A_P^s$ has the value of 1 then person $P$ uses one or more smiley in each mesage.

We introduce one more atmosphere contribution function $A_P^i(t)$ for "Interdependences" as shown eq.(4).

$$A_P^i(t) = \begin{cases} 0 & (A_Q^u(t) + A_Q^h(t) + A_Q^s(t) < 3) \\ 1 & (A_Q^u(t) + A_Q^h(t) + A_Q^s(t) \geq 3) \end{cases} \quad (4)$$

where person $Q$ is the chat partner of person $P$.
Finally, We formulate atmosphere function $A_P(t)$ as eq.(5).

$$A_P(t) = A_P^u(t) + A_P^h(t) + A_P^s(t) + A_P^i(t) \quad (5)$$

Consequently, $A_P(t)$ has a discrete value and which range is $[-1, 4]$. If $A_P(t)$ has a high value, the chat dialogue is liveliness, and if $A_P(t)$ has low value, the chat is not smooth or the person $P$ and his/her pertner are not interested in their chat topic.

### 3.2 Groups of smiley

Smiley is categorized previously into 10 groups for example *delight, grief, and amazement* and define the strength of the emotion. Table. 2 shows a part of the smiley.

Table 2: Example of smiley

| delight | (*^o^*) | (^o^) | o( )o |
|---------|---------|-------|-------|
| grief | (T_T) | (´ ) | ( `;) |
| angry | (` ´) | ( _ ´ ) | (o ´ o) |

These 10 groups consists of the 5 basic feelings frequently appered in plain text chat [7] and some behaviors such as greetings or apologizing.

## 4 Experimental results

We performed a chat experiment to get messages with smiley and to evaluate the valid of the atmosphere function.

Experimental subjects were 8 college students divided into 4 groups. We indicated each groups to chat for 10 minutes. Moreover, we asked the subjects to write down the intense of the chat dialogue atmosphere every 1 minute. Figure. 2 shows the chat winodw.

The subjects used a supporting tool to input smliley shown in Figure. 2. The subjects can register new smiley into each categories. We instructed to use this tool when they felt some kind of emotions in chat.
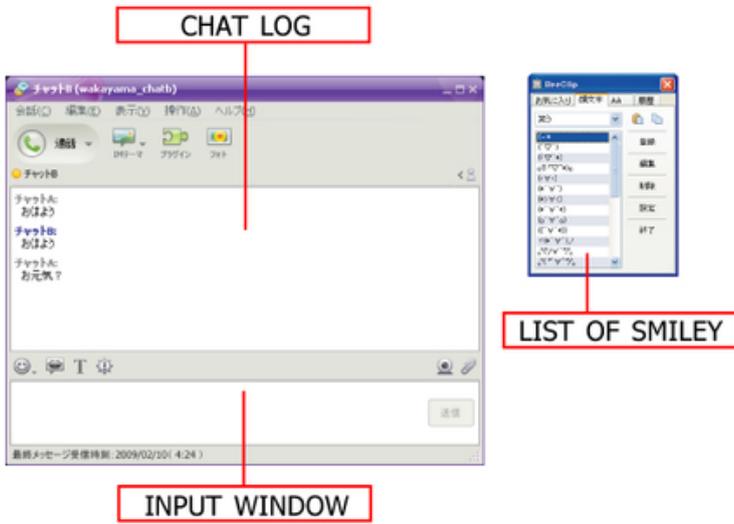
Figure 1: Chat window



Figure 2: Smiley input tool

Table 3: Average number of words, seconds for a reply and the number of smiley in all messages

| Group | Subject | Average (per message) | | Total |
| | | word (number) | time (second) | smiley (number) |
|---|---|---|---|---|
| A | a1 | 17.4 | 34.9 | 10 |
| | a2 | 25.7 | 35.9 | 9 |
| B | b1 | 16.9 | 31.1 | 3 |
| | b2 | 19.2 | 26.9 | 0 |
| C | c1 | 4.7 | 10.8 | 11 |
| | c2 | 13.4 | 17.1 | 16 |
| D | d1 | 7.7 | 20.3 | 7 |
| | d2 | 10.0 | 13.8 | 9 |



Figure 3: Graph of $A_{a1}(t)$ and $A_{a2}(t)$ in Group A

Table. 3 shows the average number of words $m_P^u$, seconds for a reply $m_P^h$, and the number of smiley in all messages $m_P^s$.

All subjects are familiar with computer. The correlation coefficient between $e_P^u(t)$ and $e_P^h(t)$ is 0.87 so the more it took time for reply, the more the subjects input words.

Figure. 3 illustrates the transitions of $A_{a1}(t)$ and $A_{a2}(t)$ in Group A. Each transition of atmosphere function are similar and the correlation coefficient is 0.48. The average of the correlation coefficient between the value of atmosphere function of person $P$ and one of person $Q$ in Group A $\sim$ D is 0.40. This result means two suites of each persons' atmosphere function tend to be similar transition, that is the subjects feel similar atmosphere from their chat. In other word, this atmosphere function can represent the positive correrations or synchronicity between two persons.

Furthermore there are some comments from test subjects that their mental state or nuance of the chat message can exchange more easily because of smiley. It was revealed that the categories of smiley are enough to communicate a user's emotion and atmosphere what he/she feel in their dialogue.

As a second step, we produced cartoons based on the chat text and the kind of smiley, to validate the definition of the atomephere function.

We produced three kind of animated cartoons from the chat log of the previous experiment. Figure. 4 is an example of cartoons. In these cartoons, two character are allocated at the bottom to make them have a conversation. The chat message log are displayed at the top of the cartoons.

The facial expressions of these characters changed according to the kind of smiley of original chat messages. A degree of the change of facial expressions is determined by the value of $A_P(t)$. When $A_P(t)$ has high level value, the character acts hammily. In contrast to this, the charater displays an expressionless face when there is no smiley in one message.

Experimental subjects were 6 college students. Four of its subjects were also subjects in the previous experiment. We asked them to watch the three cartoons and to write down the
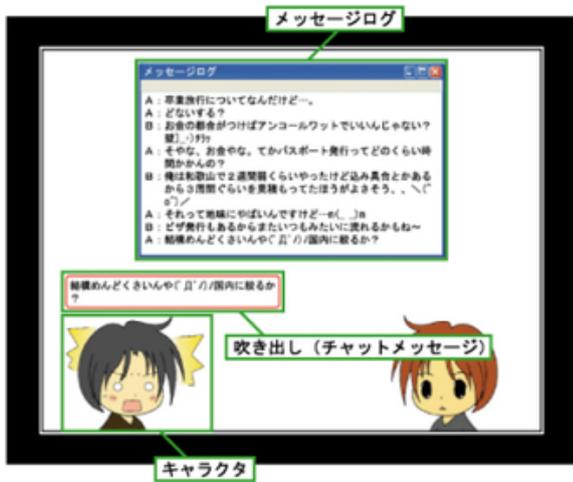
Figure 4: An example of cartoons
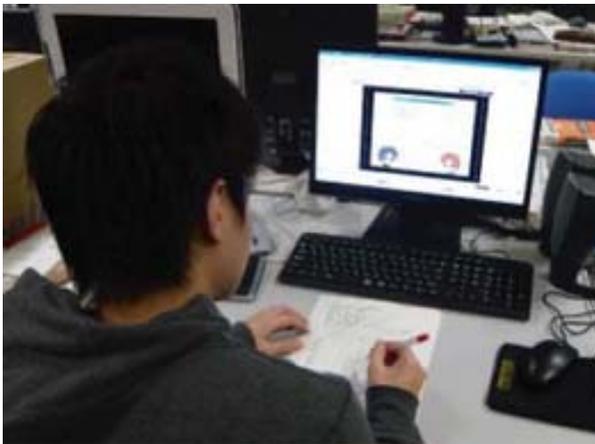


Figure 5: A scene of this experiment

atmosphsere level from $-1$ to $4$ which they feel from the cartoons. In the end of the experiment, we asked some questions. Figure. 5 shows a scene of this experiment.

We totally obtained fine rating for the three questions in Table. 4. The subjects answered that they could identify the difference between atmosphere levels and feel an equivalent atmosphere level from an original chat scene. In conclusion, the atmosphere function can express a certain level of a chat atmosphere.

However, the subjects said sometimes the atmosphere level didn't change properly in the cartoons, for instance, when the level went down. They also said it was necessary to introduce a more minus exspressions.

## 5  Conclusion

In this article, we defined the atmosphere function and investigate a relevance of this function to apply character chat system. Considering the relationships between nonverbal expressions and smliey, we represented the atmosphere function with some variables of nonverbal expressions. This function

Table 4: Questionnaire specifics

| questionnaire item | average |
|---|---|
| You can feel some atmosphere from cartoons? | 3.9 |
| These cartoons express the original scene? | 3.7 |
| The atmoephere level is represented properly? | 3.2 |

(1:no - 5:yes)

outputs high value when the number of words in one message. the time for reply if short, and a chatting person use smiley frequently.

In the experiments, we collected chat log with smiley and produced animated cartoons baesed on the chat log. In these cartoons, facial expressions of characters who played a role of agent of chat users changed according to the value of atmosphere function.

We totally obtained fine rating for the questions which are "You can feel some atmosphere from cartoons?" and "These cartoons express the original scene?".

There are some comments from test subjects that their mental state or nuance of the chat message can exchange more easily because of smiley.

Meanwhile, the atmosphere could not be expressed properly in some scene. We should investigate the cause.

## REFERENCES

[1] Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjalmsson, H. and Yan, H.: "Embodiment in Conversational Interfaces: Rea" CHI-99, pp.520-527, 1999

[2] De Carolis, B., Pelachaud, C., Poggi, I., De Rosis, F.: "Behavior Planning for a Reflexive Agent" in: Proc. of International Joint Conference on Artificial Intelligence (IJCAI 2001), pp. 1059-1066, 2001.

[3] Beattie, G. W. : "Sequential patterns of speech and gaze in dialogue" Semiotica, Vol.23, pp. 29-52 (1978).

[4] Kendon, A.: "Some functions of gaze direction in social interaction" Acta Psychologica, Vol.26, pp. 22-63 (1967).

[5] Matarazzo J.D., Saslow, G., Wiens, A.N., Weitman, M., and Allen, B.V.: "Interviewer Head Nodding and Interviewee Speech Durations" Psychotberapy:Theory, Research and Practice, Vol.1, pp.54-63 (1964).

[6] Dimberg U.: "Facial Reactions to Facial Expressions" Psychophysiology, No.6, Vol.19, pp.643-647 (1982).

[7] Itou, J. and Munemori, J.: "Analysis of Relationships between Smiley and Atmosphere during Chat with Emotional Characters" Knowledge-Based Intelligent Information and Engineering Systems: KES2008, Part III (LNAI 5179), pp.201-209 (2008).

# Practice of Installation by Apache System and learning Perl Language by Group Work

Kouji Yoshida[†], Yosuke Ando[†], Isao Miyazi[††],Kunihiro Yamada [†††]

[†]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
yoshidak@info.shonan-it.ac.jp, a063124@sit.shonan-it.ac.jp, miyaji@mis.ous.ac.jp,
yamadaku@tokai.ac.jp

*Abstract* – We developed a self-study system using Perl language. Students of department seminars and graduate schools must understand tasks ranging from installation of Apache and Perl to maintenance of CGI and servers. Furthermore, students' progress by distance learning is initially slow. We therefore support their study using a distance learning system including blended learning.

Recently, students show difficulties working cooperatively and sharing programs by work jointly. Finishing programs independently is difficult. Therefore, after using a self-study system for teaching students, we analyzed practical problems and effects of group learning. Results revealed that distance learning facilitates knowledge acquisition, but finishing programs independently is difficult. It is extremely important to use mutual teaching and regular communication with students. When active discussion, cooperation, and differing opinions among students are present, new software ideas can take shape more effectively. Nevertheless, group collaboration can also limit a student's freedom of expression of individual ideas. We report the importance of balancing student collaboration with individuals' freedom to make their own decisions. After receiving too much instruction, students' ability is inferior in terms of self-solution. If a teacher is supportive at appropriate times, then independence and ability to find solutions can be enhanced. This report clarifies our results.

*Keywords*: Apache, Perl, Group work, e-Learning, Programming, Module combination, 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. Software development has become large-scale and complicated. Consequently, independent systems are increasingly rare. Student skills related to system design and communication are not good. We must consider learning system which exploits group communication and iteration of practice, in order to develop good quality software.

During co-operating software development, the same thing must be taught repeatedly for it to become practical knowledge. 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. When a group work technique is used with this distance learning system, it can function effectively.

## 2   PROBLEMS AND PRESENT CONDITIONS

### 2.1  Problem of remote education system

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 if the system will behave as expected, if it possible to use the system as expected, and if it is possible to access the necessary contents associated with the received lectures.
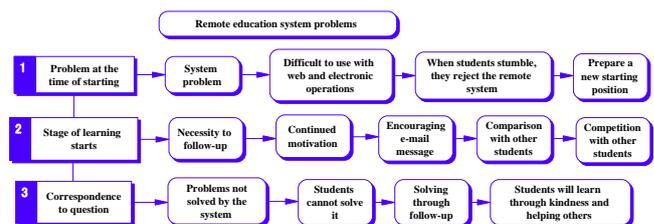


Figure 1: Problem of remote education system

### 2.2  Problem of group learning

We describe the ability, technique and knowledge by which the individual stands as displayed below. (See Fig. 2)
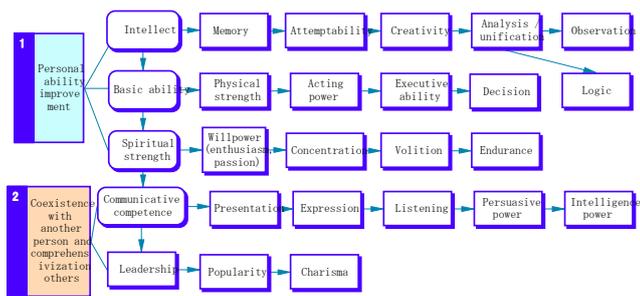
Figure 2: Ability, technology, and knowledge.

After teaching the student through lectures and seminars, we noted a particular problem to the student that we showed below. Primarily, many students cannot concentrate continuously. Concentration does not continue throughout lecturing and seminars. Some students instead surf the net, playing games that grab their interest. Secondly, many students lack self-expression and self-assertion. Moreover, when arguing with another person, students have little patience. Failure to comprehend another's opinion is an inferior ability. Thirdly, they fail to practice independence and self-restraint. They cannot make a plan and need assistance to help them determine a topic for their graduation thesis. Their tolerance for stress is poor. They are unable to bear problems and excessive pressure. Furthermore, they give up immediately and do not attend the university. To deal with these problems, it becomes important for a learning person to make positive behaviors a daily habit.

To deal with these problems, we consider a system that considers using group work technique. Three points of superiority exist by which group learning can be compared with individual learning.

(1) Students raise their intellectual power through argument among groups and learn deepened understanding.
(2) When students interact with other groups and communication, a reinforcement effect takes place, with confirmation of mutual understanding and understanding of self.
(3) When students listen to announcements of other groups, they improve their own understanding and can make a deep evaluation through the process.

## 3 SYSTEM OVERVIEW

This system comprises basic software ideas from design to programming through learning programming [1][2] . In addition, students learn applications from the basics of programming using Perl language. They learn an argument and communication through a combination and a process of a program among groups. This system has a characteristic by which learners can experience acquisition of important points when they mutually participate in a program.

This system is supported by both Linux and Windows (Microsoft Corp.) operating systems[3][4]. The student accesses the web browser via a personal computer. Students start forward learning, personal learning, and group work by

virtue of learning support and communication support. The degree of understanding is shown by solving the issue of confirmation. Learners and lecturers can communicate. Learners can examine a function of a language using a database. This System of Group Work is presented in Fig. 3.
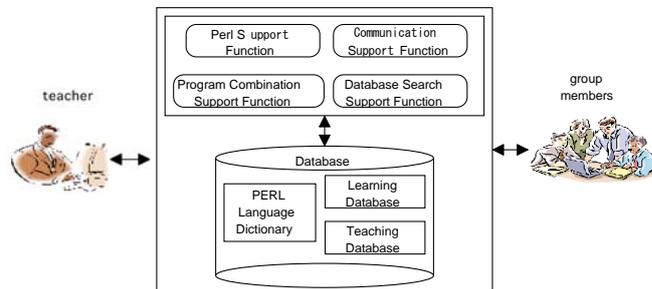


Figure 3: Group work system.

## 3.1 Support function and usage

(1) Perl language support function
When students resolve a problem of a fundamental program, the system then checks their result. Each of the students can confirm when he completes a program. When an error exists, an error message is sent using the system. Students arrive at a right answer when they revise a program using it.
· Usage of the Perl language
The program classifies items and displays 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.
· Exercise to understand The Perl Program
This exercise confirms what contents were understood by the student from using the Perl explanations given. The exercise is selective and provides a percentage representing the degree of comprehension.
· The Perl Program Exercise
This displays both the Perl program's mock validation exercise and the implementation section. Step 1 consists of validation. 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, which creates a simple explanation that is sufficiently clear for a novice programmer to comprehend. Step 2 consists of an implementation in an exercise form. They can experience a mock implementation and debugging of a complicated program. These ideas are displayed in Fig. 4.
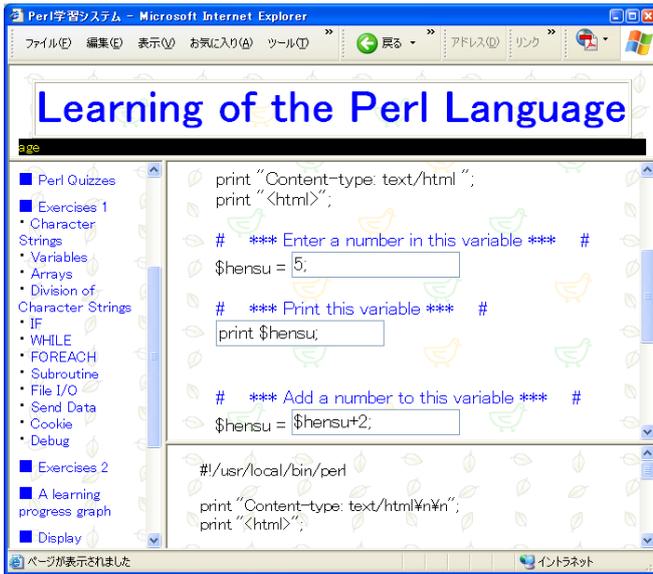
Figure 4: Supporting page of Perl language.

(2) Communication support function

· Discussion Board Q&A

This section is used for the teacher and students to exchange questions and information each other. This section also enables the teacher to respond or create messages. The background color of the page is changed when a new message is arrived in order to alert the students.

· 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 with a user-defined nickname.

· Mail transmission function and data exchange[5][6]

· Mail receiver function

· Mail transmission and history function

· Learning schedule setting and communication function

(3) Instructor support

As pointed out in the previous section describing the problem, it is important to support the learner at the initial stage. Additionally, maintaining the learner's high motivation is important. Some way must be sought to establish good communication among learners and the instructor. Especially when the learner fails to understand the received lecture, the learner must be supported with inquiries about the matter. If not, a learner might lose the desire to continue learning later. For that and other reasons, it is necessary to provide a function for the learner to examine all questions on an electronic bulletin board, for example, and to search for any one of them when necessary. In addition, it would be necessary to support learners who are not good at sending questions to the bulletin board by providing a function enabling them to communicate with the instructor individually using e-mail, for example.

On the other hand, for work such as installing a piece of software like Apache, we might encounter such a problem that merely acquiring knowledge would be insufficient to accomplish it. In such cases, real-time learner

support given from the instructor or such a blended learning system that combines the remote education system with support from the instructor would be efficient.

For example, in case the Apache installation fails to function even after checking with such notes as shown in Table 1, the next step to be provided would be a way to allow the learner to use the electronic bulletin board to discover why it does not work. A further approach would be necessary for the learner to move on to the next step to communicate with the instructor directly by asking questions in a real-time environment if the problem were not resolved after that.

Table 1: Items to check when installing Apache

| Items to verify after Apache installation | |
|---|---|
| 1 | Is Apache working properly, or is it in a state of START or STOP? |
| 2 | If Apache is working, then rewrite the config.conf file or replace it with provided one. Subsequently, START the Apache after it is STOPped once. |
| 3 | To determine whether Apache is working or not, one can try executing "index.html" of the public.html folder specifying "http://localhost/index.html" in the browser's address field. |
| 4 | If it is not working properly after going through the step described in 3 above, the installation might be incomplete, the "httpd.conf" file is not properly replaced, or it is might not have been replaced by the pre-provided one. Then, it must be re-done. |
| 5 | If an error occurred even when the pre-provided "httpd.conf" was given, the learner needs to check whether the apollo, public_html folder is created or not. If not, it must be created. |

## 3.2 Support for 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 further descriptions for practice.

(1) Practice of maintaining data integrity

This is provided to the learner to learn to show points of caution when exchanging data between 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 consistent variable names must be used 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 on 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 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 is intentionally missing an important portion as to exchange data when linking modules. Then the learner is expected to complete such an incomplete program, thereby learning the importance of data structures taken when modules are linked.

# 4 GROUP WORK PRACTICE

## 4.1 Period and method

We used a seminar method and a distance learning system for third-year students of our department[7][8]. 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 not sufficient. 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, on the premise of understanding 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 next month, students learn the basics of Perl language to file access. Then they understand the entire Perl language. They learn basic information about connections among programs using a program combination support function[9].

Students started group work seriously in the third month. First, they chose a leader among the members. Next, they argued 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.

## 4.2 Contents and results

After they argued about which program to make among 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 user interface. Students designed the data structure 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; (2) An electronic shopping lacing braid purchasing system; (3) A Web page that combined touch typing with a game".

All groups confronted difficulties related to input and output of a file 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 by this practice as well.

Figure 5 presents a bulletin board screen sample.
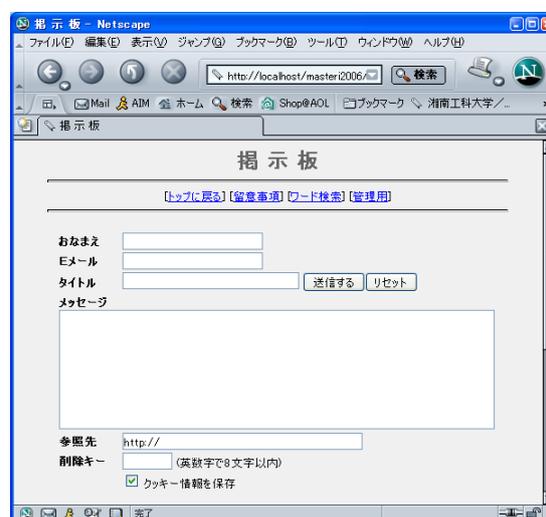


Figure 5: Sample screen for an e-Shopping Mall purchasing system

# 5 QUESTIONNAIRE RESULT AND EVALUATION

· Questionnaire contents

The learning support ability of Perl was evaluated. It differed according to the students' programming level. Students showed difficulty understanding some complex examples. It is a reference with documentation and standardization. Communication via the internet is effective for transmission of knowledge, but it is difficult to explain the complex contents concretely. Therefore, it is necessary to explain complex ideas using not only text, but also animation. Additionally, it is difficult to convey one's feelings in words.

Students are reluctant to ask other class members to review or critique course assignments. However, students excelled when practicing consultation and communication. Total communication improved by having performed group work. Actually, Perl's structures and mechanisms helped to activate group cooperation. Students learned about

documentation before and practiced it. Therefore, their consciousness was high. The necessity of documentation seems to have been realized again at the time of group work.

This is the result that asked a question for ten students. Table 2 presents questionnaire items and their respective count results. The items were rated on a scale of 1–5. Mean ratings were calculated.

Table 2: Questionnaire items and results

| | Question item | Last result |
|---|---|---|
| Group | Helping each other？ | 4.1 |
| | Amaong members good output? | 3.4 |
| | Communication between members | 3.9 |
| | Cooperation between members | 4 |
| | Problems about members | 2.2 |
| Program | Programming | 2.8 |
| | Debugging | 2.4 |
| Ability for basics | Ability for completion | 2.9 |
| | Responsibility | 3 |
| | Concentration | 3.1 |
| | Continuation power | 2.7 |
| Documentatio Other | Documentation | 3.3 |
| | Understanding | 3.1 |
| | Planning | 2.7 |
| Communicativ competence | Did the communication improve? | 4.2 |
| | Did the presentation improve? | 3.3 |
| | Did the persuasive power | 2.7 |
| | Did the listening closely | 3.6 |
| Free opinion | | |

· Evaluation and consideration

These results are expected to raise the awareness that remarkable room for improvement exists in peer-to-peer communication. Additionally, they collaborated to produce a program through an exercise; they learned cooperation with a program of a person and difficulty when they joined. Furthermore, they understood that a document was affected greatly by the communication and connection of a program. The students themselves noted the importance of their documentation; their improved motivation was noteworthy. The course items of documentation, communication, and listening were particularly well learned.

That the learner is always given correct instruction somehow without fail when failing to understand a point is extremely important. A blended learning approach gives learners a secure feeling in addition to helping the learner to have a more aggressive feeling for learning.

Because we used both electronic media and communication, students reached completion that was mutually supplemented by different learning resources.

## 6   CONCLUSION AND POTENTIAL

Fundamentally, learning for the learner requires that a student be provided with tools to learn independently. How can a system support that goal? It is true that studying in a group using a remote learning system or some associated supporting functions would be helpful for knowledge acquisition, but it would not be sufficient to master all aspects of a topic. We recognized that group teaching and group learning, and especially discussion with the group, were very efficient. We also noted, however, that such electronic communication increases problems of miscommunication from an emotional perspective. However, that problem was resolved through face-to-face communication that was available in a seminar held twice a week.

Further work in this area will develop a system that can support systemic learning by addressing electronically caused communication gaps and which can fix a group-working framework appropriately. Accomplishing those goals demands an appropriate system support that helps learners to think independently. The author plans to develop such a system to support such instruction as described, in addition to a system to support instructors as they produce guidelines and advice that are available for use in more efficient ways than ever.

## REFERENCES

[1] Linda Jorn, Ann Hill Duin, and Billie Wahlstrom, "Designing and Managing Virtual Learning Communities," IEEE Transactions on Professional Communication Vol. 39, No. 4, pp. 183–191, (1996).

[2] Raymond A. Dumont, "Teaching and Learning in Cyberspace," IEEE Transactions on Professional Communication Vol. 39, No. 4, pp. 192–204, (1996).

[3] 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).

[4] NSW Department of Education and Training, "Blended Learning,"
http://www.schools.nsw.edu.au/learning/yrk12focusareas/learntech/blended/index.php

[5] Tomohiro Nishida and Nobuaki Tokura, "The Communication between Students and Teachers via E-mail and WWW," IEICE, Vol. 96, No. 431, pp. 1–8, (1996).

[6] Kouji Yoshida, Kouiti Matsumoto, Kazuhiro Nakada, Tomonori Akutsu, Satoru Fujii and Hiroshi Ichimura, "A Trial of a Bidirectional Learning Management Tool for Promoting Learning by Mobile Phone," KES'2003, Proceedings, Part 2, Knowledge-Based Intelligent Information Engineering Systems Part 2, pp. 756–763 (2003).

[7] Jun Muenori, Hajime Yoshida, Takaya Yuizono and Masaru Sudo, "Remote Seminar Support System and Its Application and Estimation to a Seminar via Internet," IPSJ, Vol. 39, No. 2, pp. 447–457, (1998).

[8] Lisa Neal, "Virtual Classrooms and Communities," In Proc. ACM GROUP'97 Conference, November 16–19, 1997, Phoenix, AZ.

[9] Natalia Lopez, Manuel Nunez, Ismael Rodriguez, and Fernando Rubio, "Including Malicious Agents into a Collaborative Learning Envionment," ITS2002 LNCS 2363, pp. 51–60, (2002).

# Review of university educational environment learnt from "Re-study course for returning to society"

Teruhisa Ichikawa*，Ryuko Sato**

*Faculty of Informatics, Shizuoka University, Japan
**Educational Development Center, Shizuoka University, Japan

***Abstract*** - Any university has produced a lot of repeaters and those who withdraw from school. The hint of the measures was obtained from the re-study course for returning to society.

The purpose of this paper is to show a new role of the problem of the university education, the directionality of the improvement, and the university by clarifying the university education and the difference of the educational environment of this course. The main difference is many kinds of students it, classes of about ten people, immersion programs, and flexible curriculums.

Moreover, "Importance of the connection outside the university" and "Importance of the safety net that targeted the individual directly" were actually felt. It is necessary to offer opportunities for education to people the difference from the student and a variety of as a new role of the university.

***Keywords***: Re-study course, Returning to society, Educational improvement, Academe's role

## 1    Introduction

Any university has produced a lot of repeaters and those who withdraw from school. The hint of the measures was obtained from the re-study course for returning to society.

Student who withdrew from school, dropped person from society and the person who had become staying indoors attended this course. They were able to be reproduced by the course of three months.

It is a purpose of this paper to show the directionality of the problem and the improvement of the university education by clarifying the difference of the educational environment between university and this course.

## 2    Outline of re-study course for returning to society

The Ministry of Education, Culture, Sports, Science and Technology has consigned "Educational promotion business for re-study of people of society" to the university [1]. This course is one of the courses executing at 50 places or more of the university from 2007 fiscal year.

The selection result is high adoption rate as shown in Table 1.

Table 1    The selection result [2]

|  | Number of applications | Number of adoptions | adoption rate |
|---|---|---|---|
| humanities | 126 | 55 | 43.6% |
| Sc. Eng. Agr. | 56 | 22 | 39.3% |
| Medical | 48 | 19 | 39.6% |

Table 2 shows the participant whom 22 science course target. The participant of this course is not especially limiting it if equal to the bachelor. If the person who finishes high school has had the business experience for five years or more, a lecture can be attended.

Table 2    Participant [3]

| Participant | Number of courses | Composition ratio |
|---|---|---|
| Postgraduate | 9 | 40.9% |
| Bachelor | 8 | 36.4% |
| Associate degree | 1 | 0.45% |
| no limitation | 4 | 18.2% |

The course that this paper targets is "Reeducate of practicing information system study for the information engineer promotion that the region requests" that Shizuoka University is entrusting.

This course is composed of three kinds of courses as shown in Figure 1. That is, it divides into the higher class (fall term), the middle class (winter term), and the beginner's class (spring term). Six courses were executed by present [4].
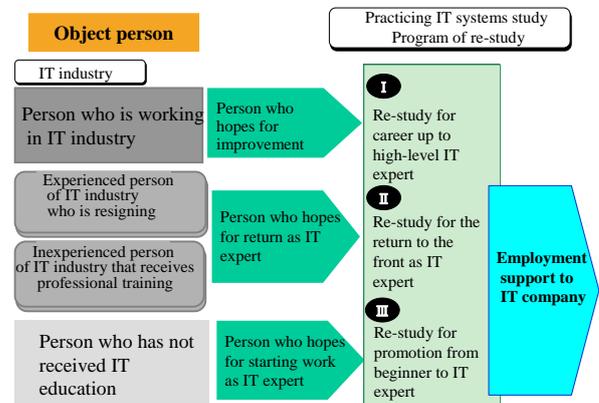


Fig. 1    Composition of re-study course

A senior course targets the information technology expert who acquired the SE experience. It is a course that does the improvement support to a high-level engineer. This course is resembled usual corporate training and removed from the object of comparison because it is not novel.

The beginner's class course targets the IT inexperienced person who hopes for starting work as an information technology expert. This course is a course (ten capacities) that offers practice and learning from the base.

Fiscal year 2008, six subjects (148 times of 90 minute) were executed from April to July.

Fiscal year 2009, the subject and the lecturer were reviewed, two subjects were abolished, one subject was newly established, and five subjects (122 times of 90 minute) were executed from April to June.



Fig. 2    Composition of beginner's class course[4]



Fig. 3    Composition of middle course[4]

A middle course targets the IT experienced person who is resigning, and aims at reemployment as an information technology expert. This course is a course (ten capacities) that supports reemployment that the studying at home is possible.

Fiscal year 2007, four subjects (73 times of 90 minute) were executed from February to March.

Fiscal year 2008, one subject was abolished, one subject was newly established, and four subjects (96 times of 90 minute) were executed from January to March.

The participant is various the ages and the experience as shown in Table 3. The route after graduating from the 3rd beginner's class course can be done about satisfactory as shown in Table 4. The route after graduating from the 5th middle course became a severe result due to the recession as shown in Table 5 .

Table 3    Breakdown of participant and graduate

| Course | Appli cation | Atten dance | Comple tion | Age | | | | Sex | | Occupation | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 20 ～ | 30 ～ | 40 ～ | 50 ～ | man | woman | Regular | Others |
| 1st | 8 | 8 | 8 | 1 | 2 | 4 | 1 | 7 | 1 | 8 | 0 |
| 2nd | 14 | 11 | 10 | 3 | 2 | 4 | 1 | 6 | 4 | 2 | 8 |
| 3rd | 19 | 13 | 10 | 6 | 4 | 0 | 0 | 6 | 4 | 0 | 10 |
| 4th | 24 | 24 | 24 | 8 | 11 | 4 | 0 | 20 | 4 | 24 | 0 |
| 5th | 13 | 10 | 10 | 5 | 4 | 1 | 0 | 7 | 3 | 0 | 10 |
| 6th | 33 | 11 | ? | 8 | 3 | 0 | 0 | 9 | 3 | 0 | 11 |

Table 4    The route after graduating the 3rd beginner's class course

| | sex | age | academic background | business experience | After graduating |
|---|---|---|---|---|---|
| 1 | male | 25 | Graduate school drop out | none | Careers (regular member) |
| 2 | male | 27 | University drop out | none | Careers (regular member) |
| 3 | female | 34 | University | 10years | Careers (regular member) |
| 4 | male | 37 | University | 9years | Independent enterprise |
| 5 | female | 28 | Junior college | 5years | Part-time job |
| 6 | male | 26 | University | 4years | Study is continued. |
| 7 | male | 34 | High school | none | Study is continued. |
| 8 | female | 36 | Special school | 6years | Part-time job |
| 9 | Female | 25 | High school | 5years | Other work is selected. |
| 10 | male | 24 | University | none | It gropes wanting do. |

Table 5    The route after graduating the 5th middle course

| | sex | age | academic background | business experience (IT experience) | After graduating |
|---|---|---|---|---|---|
| 1 | female | 32 | University | 8years (none) | Other work is selected. |
| 2 | male | 32 | University | ? (none) | Job hunting |
| 3 | male | 25 | Technical College | 1year (1year) | Careers (regular member) |
| 4 | male | 33 | Special school | 2years (2years) | Job hunting |
| 5 | female | 38 | University | 20years (4years) | Job hunting |
| 6 | Male Foreigner | 40 | Graduate school | 1.5years (1.5years) | Job hunting |
| 7 | male | 23 | Special school | 1.5years (1.5years) | Job hunting |
| 8 | male | 24 | University | 1years (none) | Job hunting |
| 9 | male | 28 | University | none (none) | Job hunting |
| 10 | female | 29 | University | 5years (5years) | Careers (regular member) |

## 3    Difference between university education and course for society persons

When the educational environment of the university is compared with the educational environment of this course, the following four points are different.

Fig. 4    Class scenery

The first difference is an environment learnt in "Various students". It learns in the student of the same generation made a round slices in the entrance exam at the university. On the contrary, it learns in various participants with different ability, age, and exposure to life in this course.

The person whom the decimal looks like very much tends to be focused on the companion formed in a homogeneous group, and many of students who drop out are solitary students who cannot be able to join a group.

On the contrary, the participant with low ability is formed the companion formed in various groups with the spirit of mutual helpfulness, feels pleasure in teaching by the participant with high ability, and feels pleasure in there is a companion who becomes kindly and teaches.

The second difference is "Fixed class of about ten people". There is a class learnt by ten people or less like the freshman seminar and the graduation research, etc. also at the university. However, one teacher only guides a specific content about once a week.

On the contrary, two or more teachers take charge of two or more classes to the participant who fixed as for this course. The sense of family bonds buds to learn by the same from morning to night companion every day.

The third difference is "Complete immersion program form". A different class is executed independently and concurrently at the university. The cooperation between subjects is not taken easily.

On the contrary, it is a subject composition in which the cooperation between subjects is required in this course. The feature is to be able to adjust the content of the succession subject according to the situation of finishing the early subject. It will be possible to learn by the immersion program efficiently in a short term. Measures to give a supplementary lecture if necessary, and not to produce those who drop out are done.

The fourth difference is "Flexible curriculum" that corresponds according to student's level and needs. The curriculum is decided by school regulations at the university, and the change is not easy. Moreover, the class improvement is being left by the charge teacher, and extent to show the total result of the student questionnaire executed every half year to the charge teacher and to refer to the improvement. Even if it is a very bad class, improvement recommendation and charge teacher's alternation powerful measures are not

taken easily.

On the contrary, the manager of the course checks the learning context of every day in this course. Opinions are exchanged with the charge teacher, and next day's educational content is reviewed if necessary. It gives a supplementary lecture if necessary. The teacher might be alternated.

## 4    Conclusions and Proposal

This course is practice of the private school (terakoya) style. The requirement to lack the university education was recognized through this education.

(1) Improvement of university education

It is not easy to change the mass production education method from the viewpoint of university management. However, it divides into the class from the 1st grader to about 10 people at Iwate prefectural university, and two or more teachers and graduate student guide this class. This method is called a stable (sumo room) style [5].

There is a fault of the private school and the stable, too. However, it is necessary to work on the improvement of the university education recognizing the goodness.

(2) New role of university

We felt the importance of the connection to the outside and the importance of the safety net through this course. The university is asked what you can do sharply as follows as giving [6].
・Increase of non-regular employment, Unstable employment increase
・Change in labor market, Changing of industrial structure
・Change in way of working, Changes in attitudes
・Hierarchical difference, The difference is social etc.

It is necessary to offer opportunities for education to people different from a present adult student.
・I want to learn at not the vocational training school but the university.
・I want you to learn even if not becoming the student and a graduate student.
・I want to taste the place and atmosphere of university.
・I want to make it to the place where integration of theory and practice.

(3) Future tasks

It is difficult to continue this business because the trust money is lost in the coming year. Therefore, it is a problem how to succeed to this knowhow.

Non-regular employment increases and the regular member is downsized   due to the recession. It is a chance to review the role that the university education plays under such an environment.

## REFERENCES

[1]    Ministry of Education, Educational promotion program for society's needs of re-study

http://www.mext.go.jp/a_menu/koutou/kaikaku/shakaiji n.htm (2009.7)

[2]  Ministry of Education, "Educational promotion program" selection situation in 2007 fiscal year http://www.mext.go.jp/b_menu/houdou/19/07/0707230 4/001.htm (2009.7)

[3]  Ministry of Education, "Educational promotion program" selection program list in 2007 fiscal year http://www.mext.go.jp/b_menu/houdou/19/07/0707230 4/002.htm(2009.7)

[4]  Let's go to the university for re-study http://www.hamanako.jp/manabinaoshi/index.html (2009.7)

[5]  Abe Y., IPU Educational System and its Effects, IEE Japan, No.1, pp.7-10 (2006)

[6]  Sato R., New role of university that considers it from the society's rechallenge by re-study , Journal of Shizuoka University Education, No.5, pp.1-8 (2009)

# Session 2: Mobile Computing
# (Chairs Yoshia Saito and Masanori Takagi)

# A Reliable Cluster-based Routing Algorithm for MANET

Hiroyuki Narumi[*], Yoh Shiraishi[**], and Osamu Takahashi[**]

[*]Graduate School of Systems Information Science, Future University Hakodate, Japan
[**]Future University Hakodate, Japan
{g2108031[*], siraisi[**], osamu[**]}@fun.ac.jp

**Abstract** - It is known that conventional reactive routing protocols in Mobile Ad-hoc Networks (MANET) decrease the packet delivery rate (PDR) in high-mobility and -traffic environments. This is because temporary congestion occurs due to control packets flooding the route establishment and the route break in long route communications. Generally, clustering is an effective method for improving the capability for dealing with node mobility. However, current clustering methods have the long route communication problem described above. Therefore, we propose a reliable clustering method, called the "Cluster-by-Cluster routing method", for reining in the emergence of long route communications by cascading short route communications. Furthermore, we implement our proposed method on Dynamic Source Routing (DSR), which is a representative of reactive routing protocol, and evaluate the effectiveness with computer simulation. As a result, in a high-mobility environment (10.0 m/s node velocity, pause time 0.0 s), our method improves PDR about 52% better than DSR and decreases the routing overhead by about 92%. In addition, in a large-scale network (250 nodes), our method improves PDR by about 22% and decreases the routing overhead by about 95%. By these results, we conclude that our proposed method can be used for reliable communication in MANET.

**Keywords**: Routing, Cluster, Large-Scale Network, Long Route Communication

## 1 INTRODUCTION

As mobile computers, such as handheld computers and tablets, continue to show improvements in convenience, memory capacity, storage capacity, and mobility, Mobile Ad hoc Networks (MANET) have become more popular. A MANET is characterized by multi-hop wireless links in the absence of any cellular infrastructure as well as by frequent host mobility. Reactive protocols, such as Dynamic Source Routing (DSR) [1], are representative protocols in MANET. We show the performance deterioration of the reactive protocol with long route communications in a high-mobility environment through basic experiments. In response to this problem, we propose a reliable Cluster-by-Cluster routing method for dividing an end-to-end long route into a combination of short routes. Furthermore, we apply our method to DSR and evaluate routing performance and scalability. The Cluster-by-Cluster routing results are compared to the results of a popular reactive routing protocol – DSR. Finally we show the effectiveness of Cluster-by-Cluster routing in high-mobility and -traffic environments.

The rest of this paper is structured as follows. Section 2 presents the problem of long route communications in MANET. Section 3 presents common clustering scheme in MANET. Section 4 presents our proposed method. Section 5 analyzes and evaluates experimental results. Section 6 concludes this paper and provides a brief outlook on our future work.

## 2 BASIC EXPERIMENT

We evaluated a reactive protocol's unicast performance in networks of varying sizes (50, 100, 150, 200, and 250 nodes) using the network simulator – version 2 (ns-2) [6]. In the simulation, we used the 802.11 standard with a transmission range of 250m. Furthermore, we chose DSR as a routing protocol, and node density of 100 nodes/km$^2$. All nodes were constantly (i.e, 0s pause time) moving according to the random waypoint model. A constant node velocity of 2.0 m/s was used. To measure the packet drop rate, we set up one constant bit-rate communication between the given source and destination node.
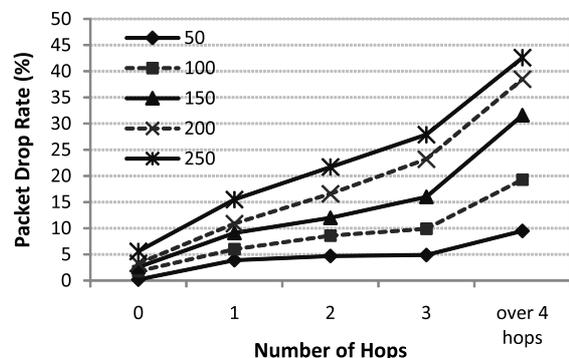


Figure 1: Packet drop rates vs. route length

Figure 1 shows the packet drop rates of the respective network sizes versus route lengths of delivered data packets. The packet drop rates in each network size increase, as route lengths get longer. For all network sizes, the packet drop rates increase quickly when route lengths are long (over 4 hops). The packet drop rate with over 4-hop communication in a large-scale network (250 nodes) increases to about 45%. As just described, long route communications in MANET cause the packet drop rate to increase. This is because extensive congestion occurs by the flooding of control packets when a source node establishes a route to long-distance destination node, and forwarding nodes drop data packets by end-to-end route breaks when they move constantly and quickly.

Consequently, it is important to reduce such flooding of control packets in the route discovery process and emergence of long route communications drop many data packets.

## 3 RELATED WORK

In this section, we present a common clustering scheme in a MANET, and describe an issue about long route communications in the scheme.

### 3.1 Clustering in MANET

To efficiently manage a network for routings (especially, the route discovery process) in large-scale and high-mobility ad hoc networks, many clustering schemes and hierarchical routing protocols commencing with the Cluster Based Routing Protocol (CBRP) have been proposed [2, 3, 4, 9, 10, 13, 14]. Figure 2 shows an example of a network formation constructed using the common clustering method used in these protocols. A cluster is a group of nodes with one of them assigned as a cluster head. A cluster is identified by its Cluster-ID, and they are either overlapping or disjointed. A cluster head will have complete information about group membership. Each node in the network has determined its corresponding cluster head(s) and, therefore, has information as to which cluster(s) it belongs to. All nodes within a cluster, except the cluster head, are called member nodes of its cluster. Any node a cluster head may use to communicate with an adjacent cluster is called a gateway node.

By dividing the network into several clusters, we are able to drastically reduce control packets called Route Requests (RREQs)[1] when a source node establishes a source route to the destination node. The route discovery process of conventional ad hoc routing protocols floods the route establishment with a large amount of control packets. However, in the hierarchical routing protocol, only cluster heads are flooded with RREQ packets in search for the source route (Figure 3). Therefore, a source node can quickly create a source route to a particular destination node. For these reasons, it is known that the hierarchical routing method has the capacity to deal with node movement compared to conventional ad hoc routing protocols such as DSR.

However, in a large-scale network, in which there are comparatively many clusters, a situation occurs where nodes physically located far from each other must communicate over two or more clusters. In such situation, the conventional hierarchical routing methods described above create long routes between source and destination nodes. As described in the previous section, if the source node sends data packets using the long route, the packet drop with the route break will increase with the common clustering method, as well as conventional ad hoc routing protocols. To deal with such packet drops caused by the movement of forwarding nodes, several hierarchical routing protocols have "Route Maintenance" and "Local Route Repair" schemes to salvage dropped packets. However, there are

---

[1] Request message of the Route Discovery process.

very few approaches for reining in the emergence of long route communications.

We adopt the concepts of an overlay clustering and our Cluster-by-Cluster routing method to reduce long route communications. Finally, we show that our Cluster-by-Cluster routing method obtains better packet delivery rates and produces lower network traffic than ever before.
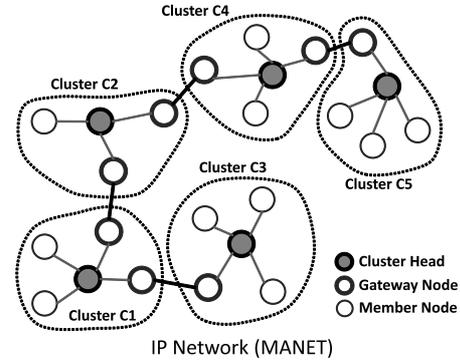


Figure 2: Clustering in MANET
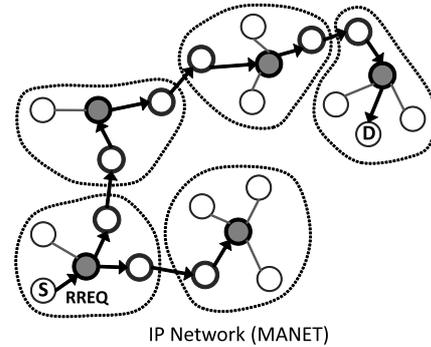


Figure 3: Route discovery process in conventional hierarchical routing method

## 4 PROPOSED METHOD

In this section, we design Cluster-by-Cluster routing method, and define a protocol layer at which Cluster-by-Cluster routing operates.

### 4.1 Overview

Cluster-by-Cluster routing forms an overlay network on a physical network, and groups physically close nodes as overlay clusters, and data packets a source node sent out are forwarded cluster by cluster. Figure 4 shows an example of this process. All nodes have complete information on the relative location of all current clusters, as described in a later section. Therefore, the source node is able to generate a proper "cluster path" in which data packets delivered on the overlay network and forwarding nodes can specify "temporal" gateway nodes currently situated at boundaries of each cluster on the path. Consequently, we can divide one physically long route communication into several physically short route communications by specifying such gateway nodes as temporal destination nodes. This leads to a reduction in long route communications, even if a source and a destination node are physically located far from each

other. Figures 4 and 5 depict examples of the route divide. Suppose source node S sends data packets to destination node D (Figure 4). Conventional hierarchical and ad hoc routing protocols create a long source route of 6 hops (Figure 5b). On the other hand, Cluster-by-Cluster routing makes 5 short source routes leveraging information about the relative location of all current clusters. In particular, a 6-hop route in conventional protocols (Figure 5b) is divided into 0-, 0-, 0-, 1-, and 1-hop short routes in Cluster-by-Cluster routing (Figure 5a). "0 hop" means a direct communication between a source and a destination node.

To dynamically construct these route divide process, we propose four core algorithms, as described below.

    1) Clustering Algorithm
    2) Acquisition of Cluster Adjacency Information
    3) Cluster Path Decision
    4) Cluster-by-Cluster Routing

We present the above four algorithms in the following sections.

## 4.2 Clustering Algorithm

We apply Random Landmarking (RLM) [5] as the base of clustering algorithm in Cluster-by-Cluster routing to map the physical topology to its overlay topology. RLM is one of the representative clustering methods for constructing P2P networks in MANETs, such as MADPastry [11, 12]. Each RLM node has a unique but dynamic overlay ID called node ID. Since there are generally no stationary works without any fixed cluster head, it uses a set of *landmark keys*. Landmark keys are simply overlay IDs that divide the overlay ID space into equal-sized segments. For example, in a hexadecimal-based ID space, an appropriate set of landmark keys could be: C000…000, C100…000, C200…000, . . . , CE00…000, CF00…000. These landmark keys are broadcasted within the corresponding cluster as part of beacon messages by cluster heads. RLM constructs clusters of physically close nodes that share a common landmark key. Therefore, physically close nodes in RLM are also quite likely to be close to each other in the overlay ID space. Figure 6 shows dynamic clustering in RLM. In our proposed method, we made two changes to RLM, "Node ID Publication" and "Node ID Lookup".
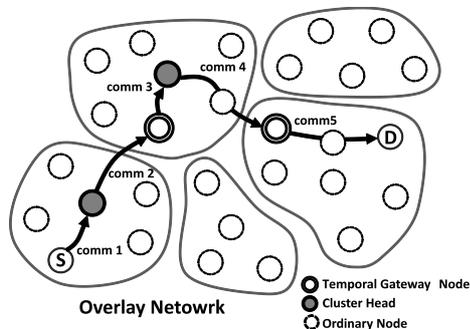


Figure 4: Cluster-by-Cluster routing
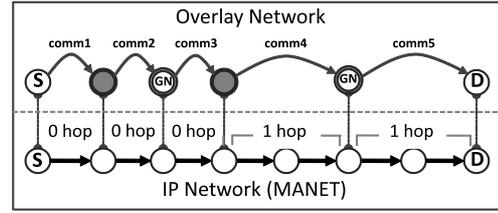


a) Route divide in Cluster-by-Cluster routing
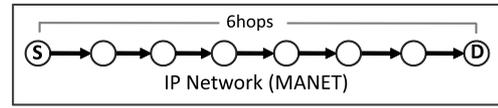


b) Long route communication in conventional routing

Figure 5: Long route communication and combination of short route communications
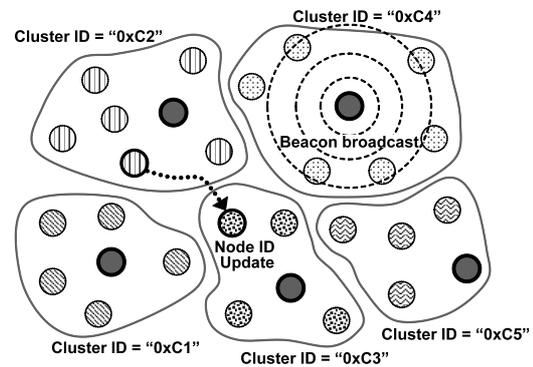


Figure 6: Clustering and node ID update in Random Landmarking

**Node ID Publication.** A node moved to a different cluster has to update its node ID to indicate its current position on the overlay network. As a result, it simply replaces its current landmark key with a new one. Then it is able to rejoin the new cluster (Figure 6). To send data packets to any destination node, a source node has to have information about the "current" node ID of the destination. Therefore, it publishes the new node ID to the node that has the numerically closest node ID to its node ID. All nodes on the overlay network obtain its current node ID, and can send data packets to it. We modified this publication process to optimize with Cluster-by-Cluster routing, as shown in Figure 7.

**Node ID Lookup.** To communicate with any node, a source node obtains that node's current node ID. Figure 8 shows the node ID lookup process in our method.



*Def inition:*
· *hash(A):* the hash value of $A$
· *generateStoreMsg(K, I, A):* Store message contains $K$, $I$, and $A$.
· $C_i$: the cluster ID of the cluster $C_i$, $0 \leqq i < n$, where $n$ is the number of clusters.

*(a) Sender:*
(1) $A \leftarrow$ my node address;   $I \leftarrow$ my node ID;
(2) $K \leftarrow hash(A)$;
(3) $M \leftarrow generateStoreMsg(K, I, A)$;
(4) send $M$ to a cluster head has cluster ID $C_i$, s.th. $| C_i - I |$ is minimal;

*(b) Receiver:*
(1) store $<K, A, I>$ pair of $M$;
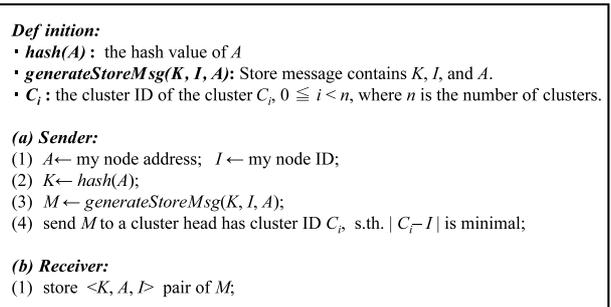
Figure 7: Node ID publication

Figure 8: Node ID lookup

## 4.3 Acquisition of Cluster Adjacency Information

To divide long route communications effectively, Cluster-by-Cluster routing collects each cluster's adjacency information, exploiting beacon messages broadcasted from cluster heads. Figure 9a shows this process. The beacon message contain a source cluster head's "cluster ID", which is called a landmark key in RLM. Then the cluster head broadcasts the beacon piggybacked with the RREQ message within its cluster. Figure 9b shows the beacon reception process. A node that received the beacon will return a "neighbor feedback" message to the source cluster head as a "temporal" gateway node if it is located at a boundary between the source cluster and the other different clusters. Otherwise, it forwards or discards the beacon as necessary. Figure 10a shows the neighbor feedback transmission process, and Figure 10b depicts the neighbor feedback reception process. Gateway nodes send the neighbor feedback piggybacked with the Route Reply (RREP) [2] message. Consequently, the cluster head that received the feedback can create a temporal bi-directional route to the gateway, the neighboring cluster. Cluster-by-Cluster routing exploits these bi-directional routes to efficiently minimize the flood of traffic during the route discovery process.

Figure 9: Beacon transmission and reception algorithm

Figure 10: Neighbor feedback transmission and reception algorithm

Figure 11 shows the acquisition process of cluster adjacency information. Suppose node A belongs to the cluster C2. If node A receives the beacon from the cluster head of C1, then A recognizes itself as the gateway node between clusters C1 and C2. Node A sets the information that C1 is next to C2 with neighbor feedback, and then returns it to C1's cluster head. C1's cluster head analyzes the feedback and updates its cluster adjacency information. After that, C1's cluster head broadcasts the beacon with the up-to-date cluster adjacency information next time, as shown in Figure 12. All cluster heads perform this operation periodically, and all nodes that receive beacons store the adjacency information each time. Finally, all nodes can obtain information about the relative location of all current clusters, as described later in Section 4.4.

**Short Route Cache.** Each node that received a beacon stores the route to the sender cluster head. In addition, each cluster head that received a neighbor feedback stores the route to the sender gateway node. This leads to effective packet forwarding and the prevention of route discovery in Cluster-by-Cluster routing.
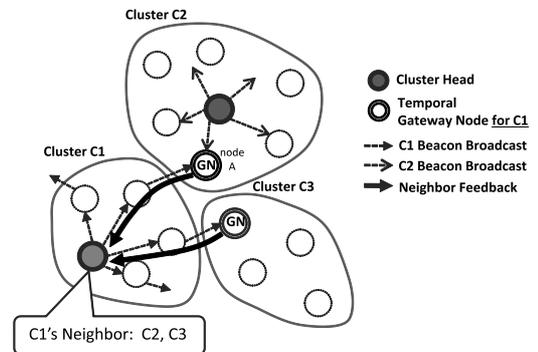


Figure 11: Acquisition of cluster adjacency information

---

[2] Reply message of the Route Discovery process.

C1's Neighbor:**C2, C3**
C2's Neighbor:**C1, C3, C4, C5**
C3's Neighbor:**C1, C2, C5**
C4's Neighbor:**C2, C5**
C5's Neighbor: **C1, C2, C3, C4**

Cluster C2
Cluster C4
Cluster C1
Cluster C3
Cluster C5
Overlay Netowrk

○ Temporal GatewayNode <u>for C2</u>
→ C2 Beacon Broadcast
➡ Neighbor Feedback
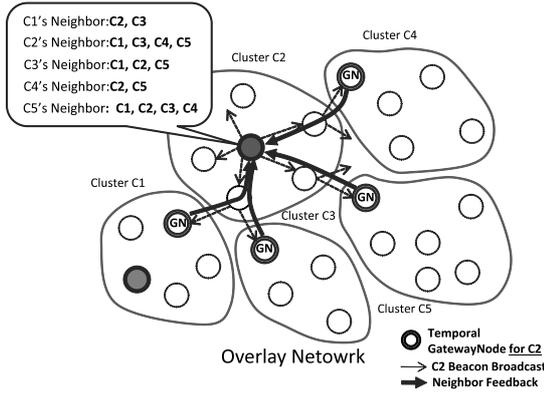
Figure 12: Acquisition of all cluster adjacency information

## 4.4 Cluster Path Decision

All nodes are able to create a cluster path to any destination cluster autonomously by exploiting its own cluster adjacency information. Before creating the path, each node generates a "cluster map", which is the relative cluster location map that base point is its cluster from the cluster adjacency information. We define clusters next to the source cluster as "Level 0", and define clusters next to Level 0 clusters as "Level 1", and so on. Figure 13 shows an example of the cluster map of cluster C0. Suppose there are nine clusters on the overlay network, and C0's cluster head collected the cluster adjacency information as shown in Figure 13a. Then nodes belonging to cluster C0 can generate the cluster map of cluster C0, as shown in Figure 13b.

By exploiting this cluster map, Cluster-by-Cluster routing creates a cluster path to any cluster the destination node belongs to. Figure 14 shows the cluster path decision algorithm. Suppose node S, belonging to cluster C0, sends data packets to any node belonging to cluster C4. From cluster C0's viewpoint, cluster C4 is a Level 1 cluster. Consequently, node S picks out Level 0 clusters next to cluster C4 from the cluster map. In this case, node S obtain two Level 0 clusters (C1 and C3), it chooses only one cluster according to any metric. Suppose node S chooses C3 as Level 0, the cluster path decision process is terminated. Therefore, node S obtains the path: "C0, C3, C4". Finally, as described in Section 4.1, we can shorten the physical route per communication using this cluster path.

## 4.5 Cluster-by-Cluster routing

A Cluster-by-Cluster routing delivers data packets cluster by cluster using the cluster path. Figure 15 shows the routing algorithm and, as mentioned above, Figure 4 depicts an example of this process. When a source node sends a data packet to a destination node, it first sends the packet to its own cluster head using a cached route obtained from the beacon. Then, the cluster head forwards it to the appropriate gateway node of the next cluster using a cached route obtained from the neighbor feedback. Cluster heads and gateway nodes of the clusters on the cluster path forward the packet until it reaches the destination cluster. Finally, the

gateway node of the destination cluster forwards it to the destination node directly.



| C6 | C7 | C8 |
| C3 | C4 | C5 |
| C0 | C1 | C2 |

C0: C1, C3
C1: C0, C3, C4, C2
C2: C1, C5
C3: C0, C1, C6, C4
C4: C1, C3, C7, C5, C6
C5: C2, C4, C7, C8
C6: C7, C3, C4
C7: C4, C5, C7, C8
C8: C7, C5

a) C0's cluster adjacency information

Cluster-Map (C0)

b) Cluster map of cluster C0
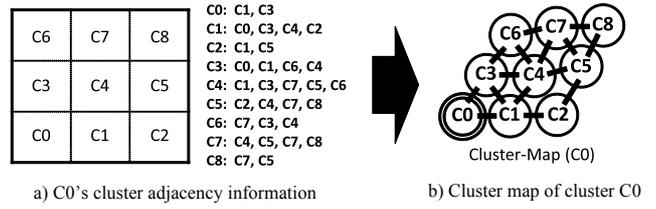
Figure 13: Cluster map generation

***Def inition:***
· ***route[ ]***: the cluster path to the destination cluster.

(1) $C_x \leftarrow$ destination cluster ID;
(2) $l \leftarrow$ level of $C_x$;
(3) $route[l] \leftarrow C_x$;
(4) while ( $l > 0$ ) {
(5)     $C_x \leftarrow$ cluster ID of level ($l - 1$) cluster that $C_x$ is adjacent cluster;
(6)     $l \leftarrow l - 1$;
(7) }
(8) // $route[]$ is a cluster path to the destination cluster
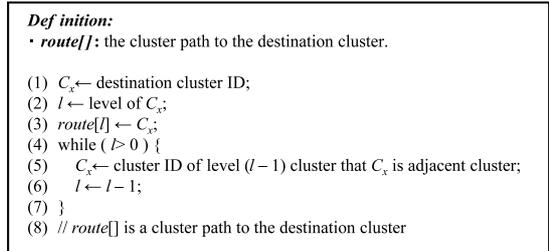
Figure 14: Cluster path decision algorithm

(1) if ( ***I am a f inal destination node*** ){
(2)     pass a packet to the upper layer;
(3) }
(4) else if ( ***I am belonging to a destination cluster*** ) {
(5)     directly forward a packet to a final destination node;
(6) }
(7) else if ( ***I am a cluster head*** ) {
(8)     forward a packet to a gateway node of next adjacent cluster;
(9) }
(10) else {
(11)     // I am a gateway or an ordinary node
(12)     forward a packet to my cluster head;
(13) }
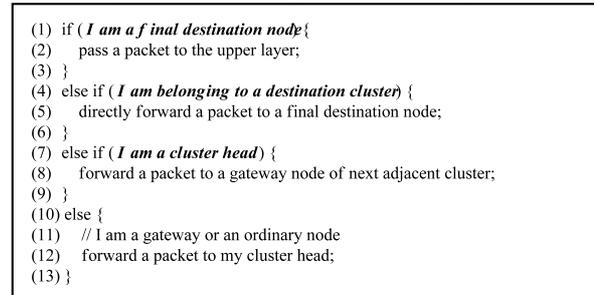
Figure 15: Cluster-by-Cluster routing algorithm

## 4.6 Layer Definition

In the TCP/IP model, the Cluster-by-Cluster routing method lies on the Internet layer, as shown in Figure 16. Cluster-by-Cluster "overhears" the surrounding packets to control the transmission of the neighbor feedback. Therefore, Cluster-by-Cluster routing operates essentially in promiscuous mode. We can use any ad hoc routing protocol supporting the promiscuous mode in Cluster-by-Cluster routing.



TCP/IP Model

Applicati on
Transport
Internet
Network Internet

Cluster−by−Cluster routing
Ad−hoc Routing Protocol
IP
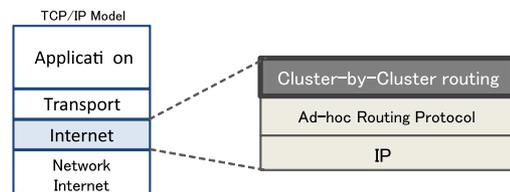
Figure 16: The TCP/IP layer model

## 5 SIMULATION RESULTS

To evaluate the performance of Cluster-by-Cluster routing, we implemented it as a routing agent in ns-2 [6]. Additionally, in this simulation, Cluster-by-Cluster routing operates on the popular reactive routing protocol - DSR.

Cluster-by-Cluster routing's results were compared to those of DSR.

## 5.1 Parameters

Table 1 lists the simulation parameters. For all simulations, we used the 802.11 standard with a transmission range of 250 m. Furthermore, we chose a node density of 100 nodes/km$^2$. All nodes were constantly (i.e. 0s pause time) moving with $V_{const}$ according to the random waypoint model, except 9 nodes allocated in a reticular pattern. That is, this network was in a high-mobility environment. Each simulation run lasted twenty simulated minutes. For Cluster-by-Cluster routing, above-mentioned 9 nodes operated as the cluster head. On the other hand, for DSR, they operate as just fixed node. The Flow State option for DSR was disabled in our simulations.

To measure the packet delivery rate and the routing overhead, two random nodes sent out a UDP message to a random node every second between 50 and 1150 seconds.

## 5.2 Evaluation Metrics

We defined three metrics to evaluate the performance of both routing agents.

**Packet Delivery Rate -** the percentage of all data packets eventually delivered to the correct destination node. This metric is calculated by the following formula.

$$Packet\ Delivery\ Ratio\ (\%) = \frac{\#\ of\ Packets\ Received}{\#\ of\ Packets\ Sent} \times 100$$
(1)

**Routing Overhead -** the total network traffic in packets that is created during the twenty simulated minutes. In the case of Cluster-by-Cluster routing, this figure comprises all route-level packets that are created by a Cluster-by-Cluster routing node: publications, lookups, beacons, and neighbor feedback. This metric is calculated by the following formula.

$$Routing\ Overhead = (\#\ of\ Control\ Packets\ Sent)$$
$$+ (\#\ of\ Control\ Packets\ Forwarded)$$
(2)

**Average Number of Hops -** the average length of the route used by data packet transmission. In the case of Cluster-by-Cluster routing, this figure comprises messages created by a Cluster-by-Cluster routing node: publications and lookups.

Table 1: Simulation parameters

| Network Parameters | |
|---|---|
| Node Density | 100nodes/km2 |
| Number of Nodes | 50, 100, 150, 200, 250 |
| Simulation Time | 1200secs |
| Node Velocity $V_{const}$ | 1.4, 2.5, 5.0, 7.5, 10.0 (m/s) |
| Node Movement Model | Random Waypoint |
| Pause Time | 0 sec |
| Transmission Range | 250m |
| Bandwidth | 2Mbps |
| **Parameters of DSR** | |
| dsragent_salvage_with_cache | true |
| dsragent_use_tap | true |
| dsragent_ring_zero_search | true |
| dsragent_enable_fl owstate | false |
| **Parameters of Cluster-by-Cluster routi ng** | |
| Number of Clusters | 9 |
| Beacon Transmission Period (unstable state) | 2secs |
| Beacon Transmission Period (stable state) | 30secs（1.4−7.5m/s）, 20secs（10.0m/s） |
| BeaconTTL（$k$） | 1 |
| Maximum Number of Neighbor Feedback（$p$） | 3 |

## 5.3 Node Mobility

Figure 17 shows the packet delivery rates that Cluster-by-Cluster routing and DSR obtain with varying node velocities (1.4, 2.5, 5.0, 7.5, and 10.0 m/s). A network size of 100 nodes was used. Cluster-by-Cluster routing obtains packet delivery rates of above 97% for all node velocities. On the other hand, the packet delivery rates of DSR drop quickly to about 42% as the node velocity increases. As a result, Cluster-by-Cluster routing's packet delivery rate versus node velocity is at most about 52% better than that of DSR with 10.0 m/s node velocity. Furthermore, as shown in Figure 18, Cluster-by-Cluster routing decreases the routing overhead about 92% more than that of DSR.

There are two reasons for DSR's markedly lower success rates.

(i) Emergence of congestion by the route discovery when a source node builds a long route to a destination node.

(ii) Drop of packet by node movement in a long route communication.

In this simulation, two DSR nodes send a message every second. If these nodes are far from each destination node, they create a long route by the route discovery. The long route is difficult to build in a high-mobility environment. If the source node fails to create the route, it retries the route discovery. This leads to extensive congestion. Once congestion occurs, many nodes cannot create their route to the destination. As a result, they discard their own messages, as their route discoveries are timed out. Even if the source node can create a long route, forwarding nodes may drop the message due to their movement. Thus, a long route communication in a high-mobility and -traffic environment is a threat to the communication reliability of the entire network. Figure 18 backs up this analysis. Figure 18 shows the routing overhead and the average number of hops versus node velocity. As the node velocity increases, the routing overhead of DSR increases drastically. This is because a large amount of congestion occurs due to the increase in long route communications, as shown in the average number

of hops in Figure 18. From the increment of the overhead, it seems that congestion occurs in 5.0, 7.5, and 10.0 m/s.

In contrast, Cluster-by-Cluster routing uses only cached short routes that are likely up-to-date, as shown in Figure 18. Therefore, there is little overhead by the route discovery in a high-mobility network.

As shown in these results, Cluster-by-Cluster routing has the capacity to deal with node movement.

## 5.4  Network Size

Figure 19 shows the packet delivery rates that Cluster-by-Cluster routing and DSR obtain in networks of varying size (50, 100, 150, 200, and 250 nodes). A constant node velocity of 1.4 m/s was used. The packet delivery rates of DSR drop gradually to about 70% as the network size increases. On the other hand, Cluster-by-Cluster routing obtains packet delivery rates of above 92% for all network sizes. As a result, Cluster-by-Cluster routing's packet delivery rate versus network size is at most about 22% better than that of DSR in the networks with 250 nodes. Figure 20 shows the routing overhead and the average number of hops versus the network size. In the network of 250 nodes, Cluster-by-Cluster routing decreases the routing overhead about 95% more than that of DSR.

The reason that packet delivery rates of DSR drop gradually with the increase in network size is the same as reason (i) in the previous subsection. In the large networks (especially 200 and 250 nodes), many DSR nodes relate to the route discovery process of one source node. Therefore, they flood the entire network with many RREQ messages. As a result, the congestion caused the deterioration of the packet delivery rate.

In contrast, Cluster-by-Cluster routing can control the emergence of long route communications. Therefore, there is also little overhead by the route discovery in the large networks. That is, Cluster-by-Cluster routing can achieve stable communications in large-scale networks because Cluster-by-Cluster routing does not use network resource.

As shown in these results, Cluster-by-Cluster routing has the capacity to deal with large-scale networks.



Figure 19: Routing overhead versus network size



Figure 17: Packet delivery rates versus node velocity



Figure 20: Routing overhead versus network size

## 6  CONCLUSION

We proposed a Cluster-by-Cluster routing method for controlling the emergence of long route communications due to clustering on the overlay network. Our simulation results show that our approach outperforms a popular reactive ad hoc routing protocol (DSR) in the scenarios considered – both in terms of packet delivery rates and routing overhead.

As future work, we plan to set the algorithm to take into account the movement of cluster heads in this Cluster-by-Cluster routing method. Furthermore, we will evaluate our method in varying combination of simulation and compare



Figure 18: Routing overhead versus node velocity

other popular ad hoc routing protocols such as AODV [7] and OLSR [8].

## REFERENCES

[1] DSR (Dynamic Source Routing) : RFC4728 http://www.ietf.org/rfc/rfc4728.txt

[2] J.Y.Yu and P.H.J.Chong, "A survey of clustering schemes for mobile ad hoc networks, " IEEE Communications Surveys & Tutorials, vol.7, no.1, pp.32–48, 2005.

[3] M. Jiang, J. Li, and Y.C. Tay, "Cluster based routing protocol (CBRP)," functional specification, IETF Internet Draft, MANET working group, *draft-ietf-manet-cbrp-spec-01.txt*, Aug. 1999.

[4] K.Lee and C.Hwang, "Hierarchical Configuration Scheme for Scalable Ad Hoc Networks," In Proc. of ICCSA, LNCS 3981, pp.816-826, 2006.

[5] R.Winter, T.Zahn, and J.Schiler. "Random Landmarking in Mobile, Topology-Aware Peer-to-Peer Networks," In Proc. of FTDCS, May 2004.

[6] The Netowork Simulator version 2 (ns-2), http://www.isi.edu/nsnam/ns/

[7] AODV (Ad hoc On-Demand Distance Vector) : RFC3561, http://www.ietf.org/rfc/rfc3561.txt

[8] T. Clausen, P. Jacquet Ed. "RFC 3626: Optimized Link State Routing Protocol," Internet Engineering Task Force (IETF) RFC, http://ietf.org/rfc/rfc3626.txt

[9] T.Ohta, S.Inoue, Y.Kakuda, and K.Ishida. "An Adaptive Mutihop Clustering Scheme for Ad Hoc Networks with High Mobility," In Proc. Of IEICE, Vol.E86-A No.7, pp.1689-1697, June 2003.

[10] R.Oda, T.Ohta, and Y.Kakuda. "An Autonomous Clustering-Based Hierarchical Routing Using Decentralized Route Information Management Scheme for Mobile Ad Hoc Networks," ICDCS Workshops 2008: pp.569-574, 2008.

[11] T.Zahn and J.Schiller. "DHT-based Unicast for Mobile Ad Hoc Networks". In Proc. Of IEEE Percom. March 2006.

[12] T.Zahn and J.Schiller. "MADPastry: A DHT Substrate for Practicably Sized MANETs," In Proc. Of ASWN, June 2005.

[13] G.He, S.Zhou, and S.Liu. "A K-hop Based Clustering Algorithm in Ad Hoc Networks," In Proc. Of NAS 2008, pp.151-152, 2008.

[14] F.Garcia, J.Solano, and I.Stojmenovic, "Connectivity based k-hop clustering in wireless networks," *Telecommun. Syst.*, vol. 22, pp. 205-220, 2003.

# A Presence-detection Method using RSSI of a Bluetooth Device

Masataka Kikawa[†1], Takashi Yoshikawa[†2], Shinzo Ohkubo[†2],
Atsushi Takeshita[†2], Yoh Shiraishi[†3], and Osamu Takahashi[†3]

[†1]Graduate School of Systems Information Science, Future University Hakodate, Japan
[†2]Research Laboratories, NTT DoCoMo, Inc., Japan
[†3]Future University Hakodate, Japan
{g2109011[†1], siraisi[†3], osamu[†3]}@fun.ac.jp
{yoshikawatak, ookubosh, takeshitaa}[†2]@nttdocomo.co.jp

***Abstract*** This paper proposes a presence-detection method using a Bluetooth equipped device. This method overcomes some of the problems with existing products and technologies for detecting user presence (e.g., security threats caused by human error, high introduction cost, and low detection accuracy). The characteristics of the Bluetooth device's RSSI (Received Signal Strength Indication) value were analyzed in preliminary experiments. According to the results of these experiments, the proposed method combines processing using RSSI value threshold and handling of instantaneous variation of RSSI value. The optimum values of the two key parameters used in the combined method (i.e., threshold RSSI value: –22 dB; and judgment time: 7 s) were determined by performance evaluation tests, and the effectiveness of the proposed method was verified by simulation.

***Keywords***: Bluetooth, RSSI, Mobile devices, Information security

## 1 INTRODUCTION

Nowadays, various information-security measures to prevent the increasing number of threats to security have been tightened. One example is "presence-detection processing" to detect when the user leaves his/her personal computer. Presence-detection is an effective measure for protecting the user's privacy and preventing information leaks and computer abuse by a third party. This measure is necessary not only for campanies but also for the general public. The existing products for presence-detection use USB devises [1] and original radiowaves [2]. These products, however, suffer problems like security deterioration if a key device (authentication key or USB key) is misplaced or lost, high introduction cost, and poor detection accuracy.

The penetration rate of mobile devices has increased year by year [3]. Additionally, most people who have a mobile device also have an internet connected personal computer (PC) [4]. In other words, there is a correlation between mobile devices and PCs. "Bluetooth" is a compact, low-power-consumption, short-range radio technology [5]. These days, many mobile devices support Bluetooth connectivity [6]. Studies on presence-detection using Bluetooth include position detection systems such as Bluetooth-based indoor proximity sensing [7], a location awareness system [8], and delay calculation [9]. These position detection systems are insufficient for presence-detection because they can only detect a small region and are generally very expensive (since they need many devices). Given those drawbacks, the authors have developed a presence-detection system using Bluetooth's received signal strength indication (RSSI) that provides relatively precise detection accuracy at modest cost.

## 2 PRESENCE-DETECTION METHOD

### 2.1 Requirement

Presence-detection for a person using a PC must meet the following two requirements. Note that 95% and over of the presence-detection processing must meet the two requirements at once.

Requirement 1: When the user moves away from front of the PC
The PC must run the presence-detection system when the distance between the user and PC is from 10 to 30 m, which was assumed as the best distance range in general.

Requirement 2: When the user is sitting in front of the PC
The PC must not run the presence-detection system when the user is sitting in front of the PC, because it would affect the convenience of user. It is also important to reduce the number of erroneous decisions.

### 2.2 Methods using Bluetooth

As for presence-detection, the two methods described below are utilized: Inquiry function of Bluetooth (inquiry method) and monitoring communication link (communication link method).

**a) Inquiry method**
A "master device" is the basic device of a Bluetooth network. A "slave device" is a device connected to the master. The master can get circumjacent slave information (Bluetooth address, Bluetooth clock, etc.) when it runs the inquiry function of Bluetooth.

The inquiry method detects the user's presence by using the inquiry function. The Bluetooth device (i.e., a Bluetooth dongle) of the PC (master) gets the slave information of surrounding Bluetooth devices by running the inquiry function regularly. The PC then judges that the user is sitting in front of the PC if the acquired slave information indicates that the user's Bluetooth device is specified as an authentication key (Fig. 1). The existing product BlueLock [10] uses this inquiry method for presence-detection.
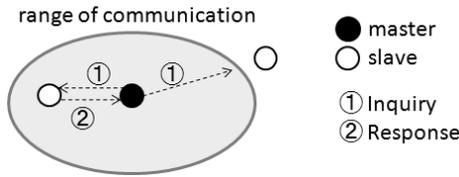
Figure 1: Inquiry method

**b) Communication link method**

A communication link is data transmission connection between a master and slave for sending and receiving data packets. The master sets up a communication link to a slave by using slave information when the master establishes a connection with the slave, and it controls data transmission to confirm whether the slave can communicate with it.

The communication link method detects the user's presence by monitoring communication links. The PC (master) monitors the communication link when the Bluetooth device connected to the PC establishes a connection with a nearby Bluetooth device as an authentication key (slave). The PC judges that the user is sitting in front of the PC if the Bluetooth device connected to the PC establishes a connection with the user's Bluetooth device specified as the authentication key(Fig. 2).
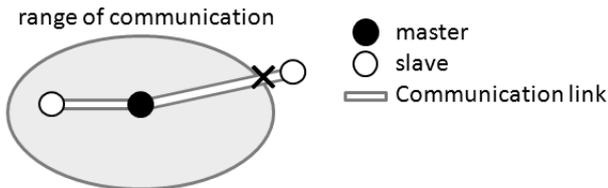


Figure 2: Communication link method

## 2.3 Performance evaluation

The success rate (pass rate) of the two methods described in section 2.2 in satisfying the two requirements stated in section 2.1 was evaluated. Fig. 3 plots the results of a performance evaluation of the inquiry method and communication link method. The commercially available product BlueLock [10] was used to evaluate the performance of the inquiry method, and a program using BlueSoleil6.2 [11] was used to evaluate the performance of the communication link method. A USB Bluetooth dongle (BT-MicroEDR2 [12]) was used as the master device, and a Bluetooth device (HT-1100 [13]) was used as the authentication key in both experiments. The USB Bluetooth dongle was placed on one side of the PC desk.

**a) Distance measurement**

Firstly, a distance measurement experiment with the two methods was performed. This experiment measured the distance between the user and the PC when the inquiry method or communication link method detected that the user had left the PC. The user walked away from the PC at a velocity of 1 m/s. In the experiment, the distance was measured 30 times by each method in three environments (namely, A: a spacious room and metal desk; B: a small room and wooded desk; C: a small room and metal desk). Fig. 3(a) plots the relationship between the requirement 1 pass rate (vertical axis) and the distance (horizontal axis) in the distance measurement experiment. The requirement 1 rate is

the proportion (in percent) of times out of 30 measurements that the method satisfies requirement 1.It is clear from the graph that either the inquiry method or the communication link method meet requirement 1 if requirement 1 pass rate is over 95%.

The inquiry method satisfies requirement 1, because the requirement 1 pass rate of BlueLock was 100%. In contrast, the communication link method does not satisfy requirement 1, because the requirement 1 pass rate of the program using BlueSoleil 6.2 was zero.

**b) Erroneous detection**

Next, an erroneous detection experiment was performed. In this experiment, the time until erroneous detection occurred while the user sat in front of the PC for 10 minutes was measured. Measurement was performed 30 times while the PC was using either of the two methods in three environments (A: spacious room and metal desk; B: small room and wooden desk; C: small room and metal desk). Fig. 3(b) plots requirement 2 pass rate (vertical axis) against time (horizontal axis) for the erroneous detection experiment. If the PC is detecting existence of the user for 10 minutes, the method will satisfy requirement 2. The pass rate in the requirement 2 is the proportion (in percent) of times out of 30 measurements that the method satisfies requirement 2. It is clear from the Fig. 3(b) that the either the inquiry method or the communication link method can meet requirement 2 if the requirement 2 pass rate is set to over 95%.

The communication link method satisfies requirement 2, because the requirement 2 pass rate for the program using BlueSoleil 6.2 was 100%. In contrast, the inquiry method does not satisfy requirement 2, because requirement 2 pass rate of BlueLock was 46.6%. These two results indicate that neither method can meet the two requirements at once.
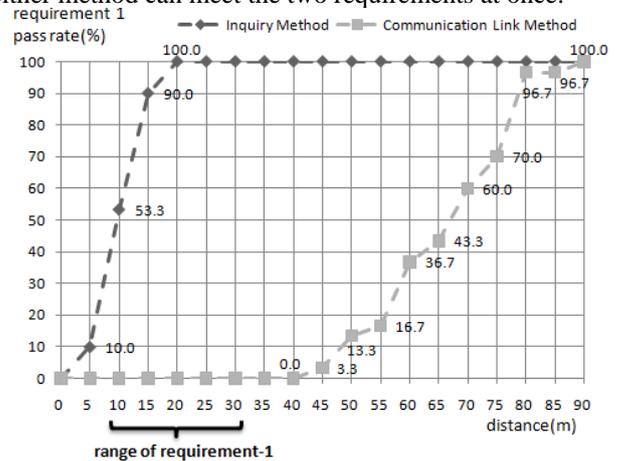


(a) Performance evaluation of distance measurement
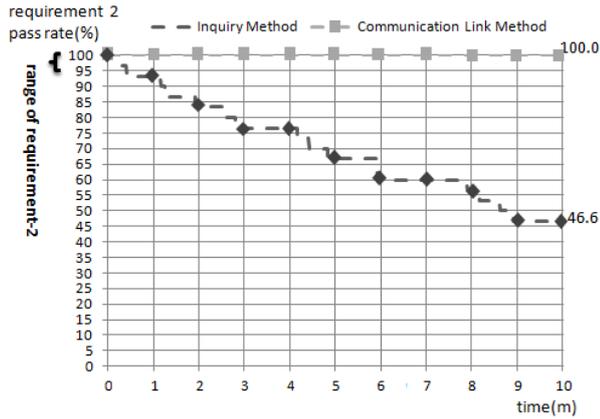
(b) Performance evaluation of erroneous detection
Figure 3: Results of evaluations of two methods

# 3 PRELIMINARY EXPERIMENTS

## 3.1 Improvement of two methods

The two methods were improved according to the results presented in section 2.3. That is, the inquiry method with improved judgment conditions was renamed the "advanced inquiry method," and the communication link method with additional comparison processing using received signal strength indication (RSSI) and a threshold value was renamed the "RSSI method".

### a) Advanced inquiry method

The advanced inquiry method uses different judgment conditions from those of the normal inquiry method, which judges that the user is sitting in front of the PC if the acquired slave information includes a Bluetooth device specified as the authentication key. Because the judgment condition used in the inquiry method is strict, this method made erroneous detections (see section 2.3) when the user was sitting in front of the PC. Accordingly, the advanced inquiry method makes the judgment in a different manner, as follows. The inquiry function is run as regularly as done by the inquiry method. If the Bluetooth function of the PC (master) gets the slave information of the authentication key once within a set period of time, the advanced inquiry method judges that the user is sitting in front of the PC. Erroneous detections can be reduced by changing the judgment conditions in this way.

### b) RSSI method

In telecommunications, RSSI is a measure of the power of a received radio signal; that is, RSSI quantifies the strength of a radio signal. In general, the value of RSSI decreases as the relative distance with a communication partner increases. RSSI can therefore be used as a measure for expressing relative distance with a communication partner.

In a similar manner to the communication link method, RSSI establishes a connection with the authentication key. The PC then begins to compare the RSSI value and the threshold value that was set for user. The PC judges that the user is sitting in front of the PC if the RSSI value is higher than the threshold value (Fig. 4). The communication link method can be finely controlled by comparing the RSSI and threshold values.

In this study, the RSSI method was chosen because it was thought that it could more effectively satisfy the two requirements than the advanced inquiry method.
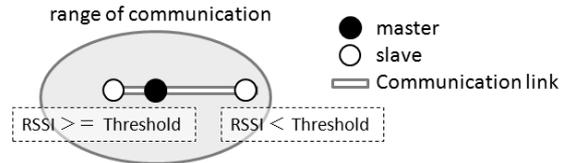


Figure 4: RSSI method.

## 3.2 Preliminary experiment on RSSI method

A preliminary experiment to determine whether the RSSI method could satisfy both requirements (Section 2.1) was performed first. It was assumed that the RSSI value changes with variations in ambient environment (e.g., desk material, position of Bluetooth dongle, and position of mobile device). In this case, the installation positions of some devices, namely, the Bluetooth dongle of the PC (i.e., the master device) and the mobile device of the user (i.e., the slave device) were different.

Table 1: Average value of RSSI

| | | Position of mobile device | | | |
|---|---|---|---|---|---|
| | | on desk | inside bag | front pocket | back pocket |
| Position of Bluetooth dongle | direct | -4.6 | -17.8 | -16.9 | -7.6 |
| | side | -9.4 | 0 | -8.8 | -0.4 |
| | floor | -21.5 | 0 | -15.0 | -18.3 |
| | reverse | -16.8 | 0 | -5.1 | -3.8 |

[dB]

Table 2: Minimum value of RSSI

| | | Position of mobile device | | | |
|---|---|---|---|---|---|
| | | on desk | inside bag | front pocket | back pocket |
| Position of Bluetooth dongle | direct | -11 | -22 | -21 | -11 |
| | side | -16 | 0 | -17 | -5 |
| | floor | -25 | -4 | -21 | -23 |
| | reverse | -23 | -2 | -14 | -9 |

[dB]

In the preliminary experiment, RSSI value was measured for three minutes when installing the Bluetooth dongle and the mobile device in various positions. Tables 1 and 2 list the average and minimum values of RSSI. The RSSI acquisition program BlueSoleil6.2 base was used to acquire RSSI values. This program can acquire the integral value of RSSI in the range of about –30 to 0 dB. If RSSI becomes less than –30, the communication link will be cut.) The Bluetooth dongle was connected to the PC at four installation positions: directly to the PC on a desk (Fig. 5: "direct"), to a USB extension cable on the right or left side of the desk (Fig. 6: "side"), on the floor underneath the desk (Fig. 7: "floor"), and on the reverse side of the desk top (Fig. 8: "reverse"). The mobile device of the user was placed in four positions: on a desk, inside a bag, in the front pocket of the user's pants, and in the back pocket of the pants.

Table 1 lists the average RSSI values for the dongle installation positions and mobile device positions stated

above. It is clear from the table that RSSI value tends to differ when the positions of the Bluetooth dongle differ. In particular, when the dongle is positioned as shown in Figs. 5, 7, and 8, the average RSSI value decreases. However, for the "side" position (Fig. 6), the RSSI value is higher than that for the other positions. The relation between RSSI (vertical axis) and distance (horizontal axis) for the side-position case when the user leaves the PC is plotted in Fig. 9(a). The relation between RSSI value (vertical axis) and time (horizontal axis) when the user is sitting in front of the PC is plotted in Fig. 9(b). As shown in Fig. 9(a), when the user leaves the PC, RSSI drops to the range of –25 to –20 dB at 10 to 30 m of requirement. Moreover, as shown in Fig. 9(b), when the user is sitting in front of the PC, RSSI momentarily drops. The position of the mobile device and operation of the user are considered to be the causes of this RSSI drop.


Figure 5: Position of Bluetooth dongle: "direct"


Figure 6: Position of Bluetooth dongle: "side"
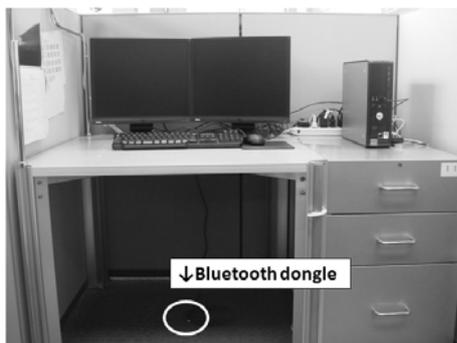

Figure 7: Position of Bluetooth dongle: "floor"
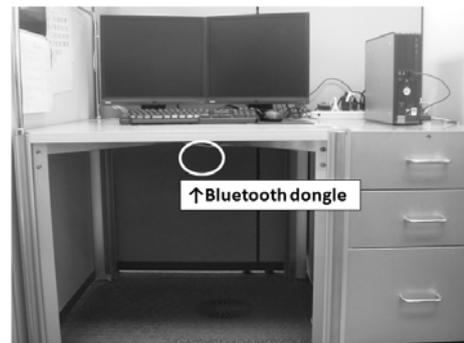

Figure 8: Position of Bluetooth dongle: "reverse"
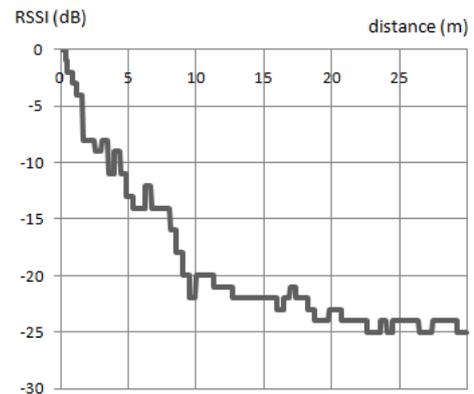

(a) When the user moves away from the PC
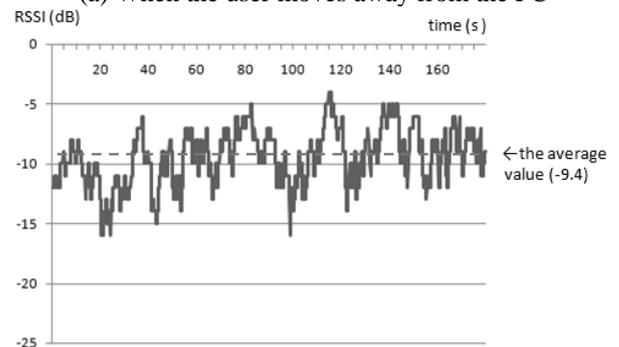

(b) When the user is setting in front of the PC
Figure 9: Examples of variation of RSSI

### 3.3 Discussion of preliminary experiment

As described in the previous section, it was shown that RSSI value changes significantly in accordance with the position of the mobile device (see Table 1) and suddenly deviates from the average value (Tables 1 and 2). When the user moves away from the desk, as shown in Fig. 9(a), RSSI decreases to about –25 dB. A high RSSI value must be maintained when the user is sitting down in front of the PC. These results indicate that the standard RSSI method (which compares a threshold value with the average RSSI value) may make an erroneous detection when the user is sitting in front of the PC.

The average and minimum RSSI values in the case that the Bluetooth dongle was connected at the desk-side installation position (Fig. 6, "side") are higher than in the cases of the other installation positions ("direct," "floor," and "reverse") regardless of the position the mobile device. The problem of varying RSSI value under different surrounding environments can therefore be solved by connecting the Bluetooth dongle at the desk-side position (because the RSSI value is fairly stable

regardless of the surrounding environment). However, since RSSI can drop momentarily, as shown in Fig. 9(b), even if the problem of RSSI-value reduction due to the surrounding environment is reduced, the RSSI method still needs to accommodate the momentary fall in RSSI value.

# 4 PROPOSED METHOD

## 4.1 Preconditions

The proposed method is an extension of the RSSI method with the following preconditions added (in accordance with the results of the preliminary experiments described above) and solves the problem of varying RSSI value due to differences in surrounding environment.

**Precondition 1: Fixation of installation position of Bluetooth dongle**

According to the results of a preliminary experiment, the desk-side installation position of the Bluetooth dongle was selected because the RSSI value is comparatively stable for this position regardless of the surrounding environment and mobile device position.

**Precondition 2: Specification of Bluetooth dongle and mobile device**

As in the preliminary experiment, a BT-MicroEDR2 (PCI) was used as the Bluetooth dongle and an HT1100 (HTC) was used as the mobile device. Fig. 10 shows the system configuration utilizing these devices.
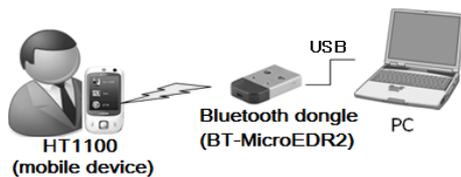


Figure 10: System configuration

## 4.2 Outline of proposed method

The proposed presence-detection method uses the RSSI function of the Bluetooth dongle to create a system that solves the problems described in Section 3.3 with high detection accuracy at modest cost. Fig. 11 shows the process flow of the proposal method, which detects the presence of a user in two processing steps using the RSSI value of a Bluetooth dongle acquired periodically.
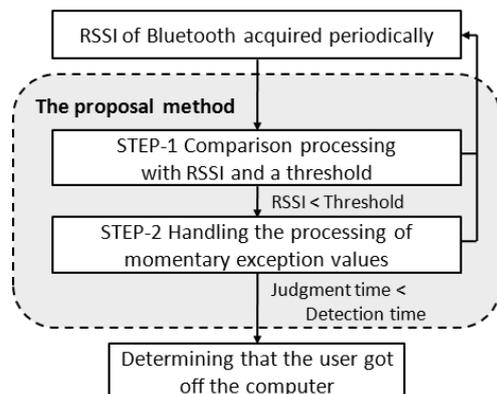


Figure 11: Processing flow of proposed method

**STEP-1** Comparison processing with RSSI value and a threshold

Fig. 12 shows an example of change of RSSI value (vertical axis) and distance (horizontal axis). The RSSI value decreases, as in Fig. 12, when the distance to a communication partner increases. Accordingly, a threshold that satisfies the requirement 1 is set. The Bluetooth dongle establishes a connection with the mobile device, and the PC compares the RSSI value of the Bluetooth dongle acquired periodically with the threshold. If the RSSI value is lower than the threshold value, the following step (STEP-2 below) is performed.
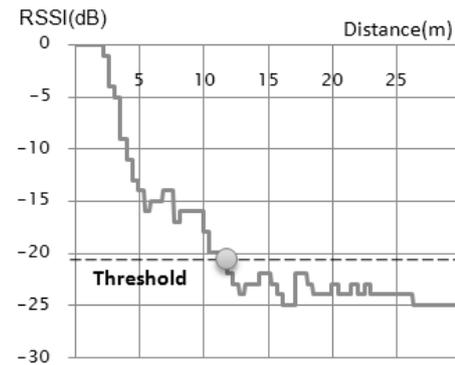


Figure 12: Example of change of RSSI with distance

**STEP-2** Handling the processing of momentary exception values

Fig. 13 plots RSSI value (vertical axis) against seating time (horizontal axis) in the case that the user is sitting in front of the PC. To satisfy requirement 2, the RSSI value must exceed a certain threshold value like that shown in data 1 in Fig 13. However, an erroneous detection might be included in the comparison processing using the RSSI value and the threshold. At this time, the PC may perform an erroneous detection because of a momentary fall of RSSI value. Accordingly, in the case of the proposed method, additional processing by using a "judgment time" as a momentary exception value is introduced.
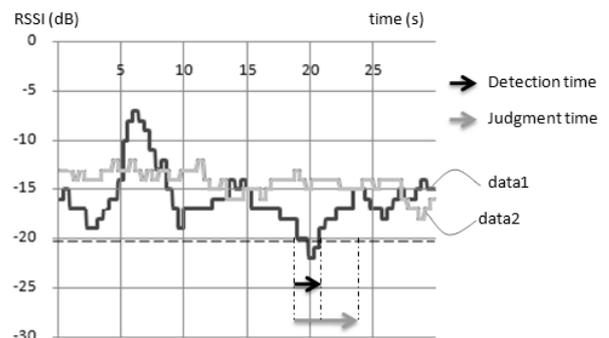


Figure 13: RSSI value against seating time in the case that the user is sitting in front of the PC.

The judgment time is set beforehand as the time that the RSSI value is monitored. In this processing, the PC measures "detection time," that is, the time that the RSSI value is less than the threshold. It then judges that the user has left the PC if the detection time is longer than the judgment time. For

example, in data 2 of Fig. 13, although the RSSI value becomes momentarily lower than the threshold, the PC judges that the user is sitting in front of the PC, because the detection time is shorter than the judgment time. Thanks to this additional processing, the proposal method reduces the number of erroneous detections.

## 5 EVALUATION EXPERIMENT

### 5.1 Evaluation data

To determine the optimum values of the threshold and judgment time and to verify the usefulness of the proposed method, the following evaluation experiment was conducted. In this evaluation, the RSSI value of the Bluetooth device was obtained during the distance measurement experiment and erroneous detection experiment (see section 2.3) performed under various five experimental environments: A spacious room and metal desk (Fig. 14); B: spacious room and wooden desk (Fig. 14); C: small room and metal desk (Fig. 15); D: small room and wooden desk (Fig. 16); E: spacious room and metal desk (Fig. 17). During the distance measurement experiment, the mobile device was in either of the following six states.

- Held in hand
- Put in front pocket of pants
- Put in back pocket of pants
- Put in a bag
- Sending a text message
- Making a telephone call

In the experiment, a total of 300 RSSI measurement were taken (i.e., five experimental environments multiplied by six mobile device states multiplied by ten times). During the erroneous detection experiment, the mobile device was in one of the following six states.

- Placed on the desk
- Put in front pocket of pants.
- Put in back pocket of pants.
- Put in bag.
- Sending a text message
- Making a telephone call

In the erroneous detection experiment, RSSI value was measured 300 times (five experiment environments multiplied by six states multiplied by ten times).
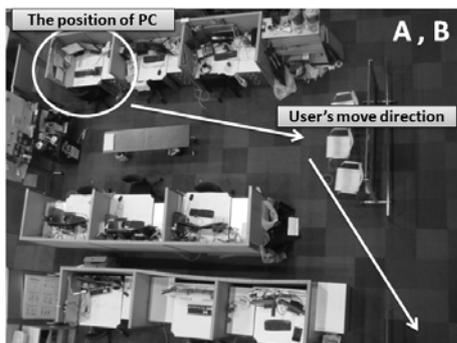

Figure 14: Experiment environments A and B
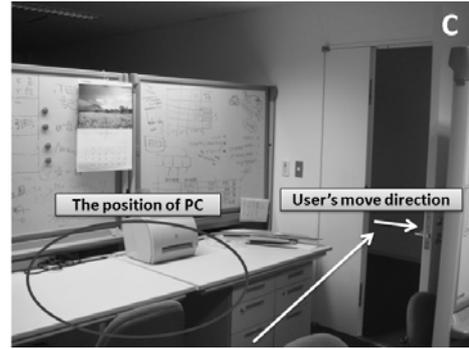

Figure 15: Experiment environment C
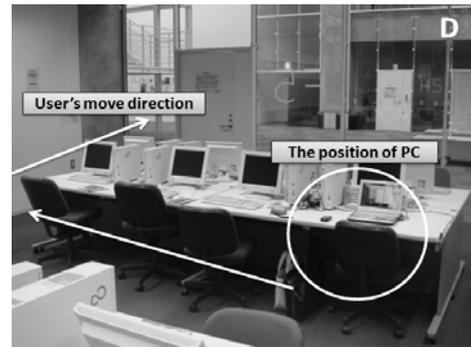

Figure 16: Experiment environment D


Figure 17: Experiment environment E

### 5.2 Evaluation simulator

The performance of the proposal method was evaluated by the simulator shown in Fig 18. This simulator outputs the result of the presence-detection by the proposed method using RSSI data (input value). It can freely set the two parameters, such as threshold RSSI value and judgment time, required for the proposal method.


Figure 18: Evaluation simulator

## 5.3 Performance evaluation

The optimal values of the two parameters (threshold RSSI value and judgment time) were determined from the evaluation data (RSSI values) given in section 5.1. Using these optimal values, the simulator (see section 5.2) was then used to evaluate the performance of the proposal method. There were thirty combinations of evaluated threshold RSSI value and judgment time as listed in Table 3.

Table 3: Combinations of RSSI threshold value and judgment time

| Th | JT | Th | JT | Th | JT | Th | JT | Th | JT |
|----|----|----|----|----|----|----|----|----|----|
| -20 | 0 | -20 | 3 | -20 | 5 | -20 | 7 | -20 | 10 |
| -21 | 0 | -21 | 3 | -21 | 5 | -21 | 7 | -21 | 10 |
| -22 | 0 | -22 | 3 | -22 | 5 | -22 | 7 | -22 | 10 |
| -23 | 0 | -23 | 3 | -23 | 5 | -23 | 7 | -23 | 10 |
| -24 | 0 | -24 | 3 | -24 | 5 | -24 | 7 | -24 | 10 |
| -25 | 0 | -25 | 3 | -25 | 5 | -25 | 7 | -25 | 10 |

Th: threshold (dB); JT: judgment time (s)

First, to evaluate whether the proposal method can satisfy requirement 1 (section 2.1), 300 measurements of the RSSI data acquired during the distance measurement experiment (section 5.1) were analyzed by the simulator. Table 4 lists the analytical results. To meet requirement 1, the PC has to perform the presence-detection processing for 95% (or more) of the time that the distance of the user to the PC from 10 to 30 m. Among the 30 combinations of the two parameters, 14 combinations meet requirement 1 (see Table 3). Moreover, when the judgment time is long, the probability of meeting requirement 1 is high.

Second, to evaluate whether the proposed method can meet the requirement 2, the 300 measurements of the RSSI data acquired in the erroneous detection experiment (section 5.1) were analyzed by the simulator. Table 4 lists the analytical results. To clear the requirement 2, the PC does not have to perform the presence-detection processing for 95% (or more)

of the time that the user is sat in front of the PC. Among the 30 combinations of the two parameters, 16 combinations meet requirement 2 (see Table 3). Moreover, when the threshold RSSI value is low, the probability of meeting the requirement 2 is high.

As shown in Table 4, when the threshold RSSI value is –22 dB and judgment time is 7 s, the proposed method meets the first and second requirements simultaneously. Accordingly, this RSSI value/judgment time combination is the optimal one under various conditions. In addition, this result demonstrates that the proposed presence-detection method is highly effective because it can acquire RSSI data in various environments (e.g., various widths of room) with difference states of mobile device. Moreover, the effectiveness of the proposed method was confirmed by simulation utilizing the results of experimental measurements.

## 6 CONCLUDING REMARKS

To develop a presence-detection system with high detection accuracy at modest cost, a presence-detection method using the received signal strength indication (RSSI) function of a Bluetooth device was developed. The performance of this method under various environmental conditions was evaluated by a simulator. According to the evaluation results, when threshold RSSI was –22 dB and judgment time was 7 s, the proposal method met the two specified requirements (section 2.1) simultaneously. It is therefore concluded that the proposal method is effective under various environments. Note that although the types of Bluetooth dongle and mobile device were specified in this study, even if other devices are used, the proposed method can be used by adjusting these two parameters (threshold RSSI value and judgment time).

In future work, the mechanism of automatically setting the parameters used in the proposal method will be investigated, automatic verification of the PC by using a mobile device will be tested, and various applications of presence-detection using the proposal method will be studied. In addition, the advanced inquiry method described in section 3.1 (another proposed method) will be experimentally evaluated.

Table 4: Results of performance evaluation of proposed method

| | | Judgment time(s) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0sec | | 3sec | | 5sec | | 7sec | | 10sec | |
| | | Req 1 | Req 2 | Req 1 | Req 2 | Req 1 | Req 2 | Req 1 | Req 2 | Req 1 | Req 2 |
| Threshold (dB) | -20 | 60.7% | **100%** | 79.0% | **100%** | 85.0% | **100%** | 88.3% | **100%** | 90.7% | **98.7%** |
| | -21 | 68.0% | **100%** | 85.3% | **99.7%** | 90.3% | **99.7%** | 92.7% | **99.0%** | 94.7% | **97.0%** |
| | -22 | 78.7% | **100%** | 90.7% | **99.3%** | 93.7% | **98.7%** | 96.0% | **95.7%** | 97.0% | 92.0% |
| | -23 | 83.0% | **99.3%** | 94.3% | **97.7%** | 96.3% | 93.3% | 98.0% | 89.0% | 98.7% | 81.7% |
| | -24 | 89.0% | 94.7% | 98.0% | 88.3% | 99.3% | 82.3% | 99.3% | 77.7% | 99.7% | 70.7% |
| | -25 | 95.0% | 87.3% | 99.3% | 70.7% | 99.3% | 64.7% | 99.3% | 60.7% | 100% | 54.0% |

Req 1: requirement 1; Req 2: requirement 2

# 7   REFERENCES

[1] Smart Signaturekey, 1 Apr. 2007, ICOMT CORPORATI
ON, 25 Jun. 2009, <http://skey.icomt.jp/>.

[2] XyLoc, 10 Oct. 2007, Nissho Electronics Corporation, 25
Jun. 2009, <http://www.nissho-ele.co.jp/product/xyloc/>.

[3] The rate of households owning a cellular phone, 20 Apr. 2
009, Yutaka HONKAWA, 25 Jun. 2009, <http://www.nis
sho-ele.co.jp/product/xyloc/>.

[4] Investigation about personal computer use of a cellular-ph
one user, 14 Nov. 2008, Japan Internet.com, 25 Jun. 2009,
<http://japan.internet.com/research/20081114/1.html>.

[5] Bluetooth, 21 Mar. 2009, Bluetooth SIG, 25 Jun. 2009, <
http://www.bluetooth.com/>.

[6] The diffusion rate of Bluetooth, 16 Sep. 2007, Hiroki YO
MOGIDA, 25 Jun. 2009, <http://techon.nikkeibp.co.jp/art
icle/NEWS/20070916/139207/>.

[7] Futoshi Naya, Haruo Noma, Ren Ohmura, and Kiyoshi K
ogure, Bluetooth-based Indoor Proximity Sensing by usin
g Device Inquiry Mechanism, IPSJ SIG Notes. ICS,Vol20
05, No.78, pp. 5–10 (2005).

[8] Takatoshi Takakai, Masahiro Fujii, and Yu Watanabe, A
Study on Location Awareness System using Cellular Pho
ne with Bluetooth, IEICE technical report. USN, Vol. 107,
No. 53, pp. 61–66 (2007).

[9] Naoki Horiuchi and Nobuo Nakajima, Indoor Positioning
by Bluetooth, Proceedings of the IEICE General Conferen
ce, Vol. 2004, No. 1, p. 827 (2004).

[10] BlueLock, 8 Sep. 2008, Wuul, 25 Jun. 2009, <http://me
mbers.lycos.co.uk/wuul/ >.

[11] BlueSoleil, 21 May. 2009, IVT Corporation, 25 Jun. 200
9, <http://www.bluesoleil.com/>.

[12] BT-MicroEDR2, 23 Nov. 2007, PLANEX, 25 Jun. 2009,
<http://www.planex.co.jp />.

[13] docomo FOMA HT1100, 12 Jun. 2008, HTC, 25 Jun. 20
09, <http://www.htc.com/jp/ >.

# QoS Evaluation for Real-time Distributed Systems Using the Probabilistic Model Checker PRISM

Takeshi NAGAOKA[†], Akihiko ITO[†], Kozo OKANO[†], and Shinji KUSUMOTO[†]

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

The Internet is a best-effort network shared by a number of users, so there is no guarantee on the QoS properties such as network bandwidth, delay and throughput. Therefore, system developers preliminarily have to estimate the QoS by simulation techniques or mathematical analysis. Probabilistic model checking can evaluate performance, dependability and stability of information processing systems with random behaviors. PRISM is one of probabilistic model checkers. It handles automata with probabilities (discrete and continuous time Markov chains) and time elapse. This paper presents a discrete Markov chain for a real-time distributed system, and evaluates its QoS properties using the simulation function of PRISM. As the target real-time distributed system, we use an experimental system whose main subsystem is video data transmission which uses RTSP (Real Time Streaming Protocol) for the streaming protocol. The system has also ftp servers and clients which exploit tcp connections, as well as a packet generator that generates udp packets as background noise. Thus, the system involves several simultaneous connections; it may occur congestion. To validate the correctness of our model, we also model it in a model for the well-known network simulator NS–2. The paper gives the comparison of their simulation results. The comparison shows that the result of PRISM simulation is very similar to that of NS–2. The results show that the proposed method is useful to analyze the network performance.

***Keywords***: Real-time Distributed System, Model Checking, PRISM, Formal Verification

## 1 Introduction

Nowadays real-time distributed systems like streaming media systems are widely spreading. These systems require time based transmission such as QoS control to prevent interruption of packet transmission caused by network delay, packet loss, and so on.

Since the Internet is a best-effort network shared by a number of users, there is no guarantee on the QoS properties such as network bandwidth, delay and throughput. Therefore, system developers preliminary have to estimate the QoS by simulation techniques[2] or mathematical analysis[4].

Simulation techniques usually do not guarantee qualitative properties such as the maximum throughput and the minimum jitter, and so on, though they can calculate mean-values along typical traces. In general, these techniques use much resources to simulate accurately the target network systems. On the other hand, mathematical analysis is logically correct and needs fewer computation resources, but in many cases the based models are too simple and ideal; hence it is hard to apply the mathematical analysis to realistic applications.

Formal verification techniques, especially model checking techniques[3] are considered as promising techniques for information system developing due to their ability of exhaustive checking. Among them, probabilistic model checking can evaluate performance, dependability and stability of information processing systems with random behaviors. PRISM[1] is one of probabilistic model checkers. It handles automata with probabilities (discrete and continuous time Markov chains) and time elapse. Therefore, it is suitable for modelling the network systems.

This paper presents a discrete Markov chain for a real-time distributed system, and evaluate its QoS properties using the simulation function and the verification function of PRISM. As the target real-time distributed system, we use an experimental system shown in Fig.1. The main subsystem of the system is video data transmission which uses RTSP (Real Time Streaming Protocol) for the streaming protocol. The system has also ftp servers and clients which exploit tcp connections, as well as a packet generator that generates udp packets as background noise. Thus, the system involves several simultaneous sessions; it may occur congestion.

This paper analyzes newly throughput property. The analysis is performed on two levels; one is detail models, and the other is abstract level. On the detail level, several numeric analysis is performed, whereas on the abstract level, qualitative verification is performed. To validate the correctness of our model, we also model it in a model for the well-known network simulator NS–2[2], and give the comparison of their simulation results. The comparison shows that the result of PRISM simulation is very similar to that of NS–2.

As related works, several case-studies are performed using PRISM [8]–[10]. For example, Paper [9] deals with network protocol.

The rest of the paper is organized as follows. Sec.2 gives some backgrounds, especially overviews of PRISM and NS–2, as preliminaries. Sec.3 describes the example real-time distributed systems which uses streaming media. Explanations on several famous protocols for real-time video systems are also given. Sec.4 and 5 provide concrete models of the example network described in Prism and NS–2, respectively. Sec.6 and 7 show the results of comparison and gives discussions. Finally Sec.8 concludes the paper.

## 2 Preliminaries

This section simply describes Prism and NS–2 as well as network protocols for net streaming.

## 2.1 Probabilistic Model Checking tool PRISM

Here, we simply describe overview of probabilistic Model Checking tool PRISM[1].

A model checking tool usually has two inputs, a model $\mathcal{M}$ and a logical expression $p$. The model is typically a transition system which represents behavior of the system to check; while the logical expression is a temporal logic expression which represents a property to check. The typical output of the model checking tool is whether the logical expression is valid on the model ($\mathcal{M} \models p$). Some model checker outputs a counter example when $p$ is invalid.

The inputs of PRISM include the following three kind of transition system as a model:

Discrete-time Markov chains (DTMC);

Continuous-time Markov chains (CTMC); and

Markov decision processes (MDP).

Each of three systems is a probabilistic transition system (Markov chain). The inputs of PRISM also include Probabilistic Computation Tree Logic (PCTL)[11] for DTMC and MDP, and Continuous Stochastic Logic (CSL)[12] for CTMC. They are CTL based logics enchanted with probability.

PRISM has several analysis modes: simulation mode, numerical analysis mode, and verification mode. Using the simulation mode, we can observe the behavior of the given model system visually. This mode uses only a model as input. Numerical analysis mode can evaluate the value of uncertain variable specified with PCTL or CSL based on the model. Such numerical analysis is considered as a kind of parametric model checking[13]. PRISM can draw a graph with several trials of such numeric analysis. Verification mode is like typical model checking except that PRISM cannot output counter examples.

In this paper, we use DTMC's as the model of the network. Here, we describe more precisely on a DTMC. Formally, a DTMC $D$ is a tuple $(S, s\_init, P, L)$, where

$S$ is a set of states ("state space") ;

$s\_init \in S$ is the initial state;

$P : S \times S \rightarrow [0, 1]$ is the transition probability matrix where $\Sigma_{s' \in S} P(s, s') = 1$ for all $s \in S$; and

$L : S \rightarrow 2^{AP}$ is function labelling states with atomic propositions.

PRISM allows a transition to specify an action and updating expressions on $D$, where $D$ is a set of variables with finite domain. In other words, a DTMC of PRISM is a kind of extended automaton with probabilities. Usually, one execution of a transition is translated into a unit time of time elapse (a tick event). Such scheme is known as digital clock view of DTMC, however, using an integer variable (with the upperbound) explicitly as a clock variable, we can also represent a system with discrete time in DTMC. In this paper, we use the latter scheme to avoid the state explosion problem.

## 2.2 Network Simulator NS–2

The Network Simulator (ns) is a network simulator developed by Virtual InterNetwork Testbed (VINT) project. It has been widely used[16].

NS–2 is the successor version of ns. It is open source software. NS–2 uses two languages: C++ for core libraries; and OTcl (Object-oriented Tcl) for scenario description and network configuration. A new protocol which is not pre-defined in the core libraries has to be written by a user.

## 2.3 Protocols for Net-streaming

Here, we simply summarize typical protocols used in the Internet. Typical protocols used in the Internet have a congestion control mechanism in order to avoid network congestion. For example, TCP (Transmission Control Protocol) uses AIMD (Additive Increase Multiplicative Decrease) type window-flow control as such the mechanism. It controls the data size of sending packets based on the current available bandwidth. Such scheme has an advantage for the correct data transmission. It, however, allows delays, which is not suitable for real-time data transmission. Therefore, RTSP (Real Time Streaming Protocol) is used for real-time application. RTSP is a protocol for the Internet streaming of voice and movies, on TCP/IP network. Famous congestion control mechanisms for RTSP are RAP (Rate Adaptation Protocol)[5] and TEAR (TCP Emulate At Receivers)[6]. Recently, TFRC (TCP-Friendly Rate Control)[7] attracts attention. Hence, this paper models TFRC.

### 2.3.1 RTSP

RTSP is one of typical protocols working at end-to-end. RTSP has five states, called SETUP, PLAY, RECORD, PAUSE and TEARDOWN. RTP (Real-time Transport Protocol) is responsible for transmission of stream-data. It determines the throughput of RTP based on rate control scheme of TFRC using the report message of RTCP (RTP Control Protocol).

### 2.3.2 TCP Friendly Rate Control TFRC

TFRC is a rate control scheme for fairness between RTP and TCP. It controls the rate in order to avoid bad effects on existing TCP flows in the same network, which increases total effectiveness of the whole network. TFRC controls the rate using the report message of RTCP. The report message contains loss of packets and jitters, which can be estimated via the sequence number of received RTP packets and time stamps, respectively. RFC3448 describes the following formula for determining the throughput:

$$X = \frac{s}{R * \sqrt{2 * b * p / 3} + (t\_RTO * (3 * \sqrt{3 * b * p / 8} * p * (1 + 32 * p^2)))}$$

where the unit of $X$ is byte/seconds. The parameter of the formula is summarized in Table 1.

The calculated throughput is a rate with which a RTSP server should send packets considering the network congestion at the time. Therefore weighted average values of the pa-
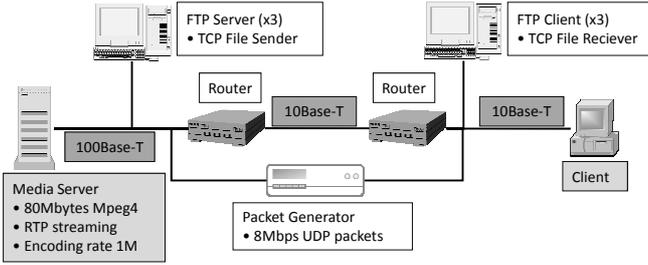
Figure 1: A Configuration of Experimental System

rameters in a short period are applied into the equation. Paper [7] also defines the calculation methods for the parameters.

When the value of $X$ is less than the bandwidth, TFRC lets RTSP set the value as throughput.

## 3  Example Network Configuration and Protocols

In this section, we introduce an example of a real-time distributed system. As the example system, we select a video data transmission system[14] shown in Fig.1. The system is composed of a pair of a video server and its client, three pairs of FTP servers and their clients, and a packet generator, which are connected to each other through routers located at the middle of Fig.1. The routers are connected through the 10Base–T Ethernet, which is considered as a bottleneck of packet transmission. In the considering scenario, the video server sends 80MB of video data with throughput of 1Mbps using the rate control of TFRC. After 100 seconds from the start of the video streaming, FTP servers and clients start their data transmission through TCP sessions. Also, the packet generator always sends UDP packets with the throughput of 8Mbps as background noise.

## 4  PRISM Model

We generate two PRISM models for the example of Fig.1. The first one is modelled in detail, and another one is based on the simulation results on the first model but is more simplified model. First, we apply PRISM simulation to the detailed model to show the correctness of our model and to obtain the data to generate the simplified model. Next, the simplified model is used to verify the qualitative properties of the throughput.

Table 1: Parameters of the Throughput Estimation Formula

| R[seconds] | Round trip time |
|---|---|
| p[%] | A packet loss rate |
| s[byte] | Packet size |
| b[number of times] | A number of packets acknowledged by a single TCP acknowledgment |
| t_RTO[seconds] | A TCP retransmission timeout value |

## 4.1  The Detailed Model

In the detailed model, we have to model packet loss rates and round trip time of the RTP packets as well as the buffer control of the routers, the congestion control of transmission protocols, and a time elapsing mechanism.

### 4.1.1  Time Elapsing

In order to manage time elapsing in the detailed model, we declare an integer variable to represent time. Time elapsing is based on events such as packet transmission. In the model, each module registers time of occurrence of the next event, and when all modules register the time, the model perform time elapsing into the latest time of the registered event from the current time. After time elapsing, a corresponding module performs the registered event and registers time of the next event again.

The module to manage time elapsing contains two variables and is implemented as ten lines of code.

### 4.1.2  Buffer Control of the Routers

Buffer control of the routers is modelled as a queue which holds transferring packets. In the model, the current queue length is managed with an integer variable. Enqueue and dequeue behaviors are described in the model as operations. Also, regardless of the current queue length, enqueueing packets are dropped with certain probability. In order to construct the module of buffer control, we have to specify maximum length of the queue, a packet transfer rate of the link between the routers, and constant probability to drop enqueueing packets as parameters.

In the enqueue operation, if the current queue length becomes larger than the maximum one, the enqueueing packet is considered to be dropped (drop tail). The dequeue operation is abstracted; together with the time elapsing operation, a number of packets are output from the queue at a time according to the packet transfer rate of the link.

The buffer control module shown in Fig.2 contains ten variables and is implemented as about 80 lines of code, where the module also manages the history of packet loss intervals used for the congestion control. In Fig.2, the ten variables are firstly declared. Several actions are defined in CCS like expressions with probabilities. For example, the expression

```
[ENQFTP1]  (q_len <= MAXQLEN - pnum_ftp1)
  -> 1 - P_LOSS_RATE : (q_len' = q_len + pnum_ftp1);
   + P_LOSS_RATE : true;
```

stands for that when an action $ENQFTP1$ occurs and ($q\_len \leq MAXQLEN - pnum\_ftp1$) holds, the variables $q\_len$ is updated to $q\_len + pnum\_ftp1$ with probability $1 - P\_LOSS\_RATE$, or do nothing with probability $P\_LOSS\_RATE$.

### 4.1.3  Congestion Control

In the model, TCP and RTP sessions can handle congestion control.

The slow-start and congestion avoidance behaviors of TCP are embedded to our TCP model. For each connection of the TCP, we declare two integer variables to manage the slow-start threshold and the window size. For each connection of

```
module Router

  q_len : [0..MAXQSIZE] init 0; //The current queue length

  //The history of packet loss intervals
  int_p_loss0 : [0..10000] init 0;
  int_p_loss1 : [0..10000] init 10000;
  int_p_loss2 : [0..10000] init 10000;
  int_p_loss3 : [0..10000] init 10000;
  int_p_loss4 : [0..10000] init 10000;
  int_p_loss5 : [0..10000] init 10000;
  int_p_loss6 : [0..10000] init 10000;
  int_p_loss7 : [0..10000] init 10000;
  int_p_loss8 : [0..10000] init 10000;

  //A flag to observe whether the packet loss occurs burstly or not
  p_loss_flag : bool init false;

  //When the queue length does not reach to its maximum
  //Transfering packets are dropped with certain probability
  [ENQMS] (q_len <= MAXQLEN - ms_pnum)
    -> 1 - P_LOSS_RATE :
       (q_len' = q_len + ms_pnum) & (p_loss_flag' = false) &
       (int_p_loss0' = int_p_loss0 + 1)
     + P_LOSS_RATE :
       (int_p_loss0' = 0) & (int_p_loss1' = int_p_loss0) &
       (int_p_loss2' = int_p_loss1)&(int_p_loss3' = int_p_loss2)&
       (int_p_loss4' = int_p_loss3)&(int_p_loss5' = int_p_loss4)&
       (int_p_loss6' = int_p_loss5)&(int_p_loss7' = int_p_loss6)&
       (int_p_loss8' = int_p_loss7);

  //When the queue length reaches to its maximum
  [ENQMS] (q_len > MAXQLEN - ms_pnum) & (!p_loss_flag)
    -> (q_len' = MAXQLEN) & (int_p_loss0' = 0) &
       (int_p_loss1'=int_p_loss0+1)&(int_p_loss2' = int_p_loss1)&
       (int_p_loss3' = int_p_loss2)&(int_p_loss4' = int_p_loss3)&
       (int_p_loss5' = int_p_loss4)&(int_p_loss6' = int_p_loss5)&
       (int_p_loss7' = int_p_loss6)&(int_p_loss8' = int_p_loss7)&
       (p_loss_flag' = true);

  // When the packet loss occurs burstly
  //      (do not update the history of packet loss intervals)
  [ENQMS] (q_len > MAXQLEN - ms_pnum) & (p_loss_flag)
    -> (q_len' = MAXQLEN) & (p_loss_flag' = (q_len = MAXQLEN));


  //The ENQUEUE operations for the three FTP sessions
  [ENQFTP1]  (q_len <= MAXQLEN - pnum_ftp1)
    -> 1 - P_LOSS_RATE : (q_len' = q_len + pnum_ftp1);
     + P_LOSS_RATE : true;

  [ENQFTP1]  (q_len > MAXQLEN - pnum_ftp1)
    -> (q_len' = MAXQLEN);

  [ENQFTP2] (q_len <= MAXQLEN - pnum_ftp2)
    -> 1 - P_LOSS_RATE : (q_len' = q_len + pnum_ftp2)
     + P_LOSS_RATE : true;

  [ENQFTP2]  (q_len > MAXQLEN - pnum_ftp2)
    -> (q_len' = MAXQLEN);

  [ENQFTP3]  (q_len <= MAXQLEN - pnum_ftp3)
    -> 1 - P_LOSS_RATE : (q_len' = q_len + pnum_ftp3)
     + P_LOSS_RATE : true;

  [ENQFTP3]  (q_len > MAXQLEN - pnum_ftp3)
    -> (q_len' = MAXQLEN);

  // DEQUEUE is executed together with the time elapsing event
  [TIMER] (q_len != 0)
    -> (q_len' = max(0, q_len - floor(min_lookahead *
                         (PACKETPERMSEC -PACKETPERMSEC2))));

  [TIMER] (q_len = 0) -> true;

endmodule
```

Figure 2: The Module of Router Described with PRISM Language

the TCP, we declare six variables and the behavior is implemented as about 30 lines of code.

In the transmission of RTP, a packet transmission rate is calculated from the throughput equation defined in [7]. To use the equation, we also have to model a packet loss rate and round trip time of the RTP session (see Sec.4.1.4 and 4.1.5). The module related to RTP contains six variables and is describe with about 40 lines of code.

In our model, the packet transmission behaviors of TCP and RTP are abstracted as a number of packets are transmitted simultaneously.

### 4.1.4 Round Trip Time

In our model, round trip time is obtained using physical delay and delay in the router. The delay in the router is calculated as the time to transmit all packets currently buffered in the router. Therefore, we obtain the delay in the router using current queue length and a packet transmission rate of the link.

### 4.1.5 Packet Loss Rate

In the TFRC specification[7], a packet loss rate is calculated using intervals of packet loss. To avoid the loss rate varying rapidly, a history of the packet loss intervals is used to calculate it. Our model declares eight integer variables to manage the history. In the calculation of the loss rate, recent intervals in the history are weighted heavily.

## 4.2 The Simplified Model

The detailed model described in Sec.4.1 is too complicated to verify its qualitative properties using model checking. Here, we create a simplified model based on simulation results on the detailed model to perform model checking. Using the simplified model, we verify the minimum and maximum throughput of the media server.

The simplified model is based on the throughput equation defined in [7]. In order to create the simplified model, we have to measure distributions of round trip time and packet loss rates on the detailed model. From the obtained distributions, we create discrete probability distributions of them. The simplified model decides its round trip time and a packet loss rate according to the probability distributions. Finally, by verifying occurrence probability of the throughput on the simplified model, we can obtain the minimum and maximum throughput.

The simplified model is implemented as 50 lines of code.

## 5 NS–2 Model

We use a simulation scenario shown in Fig.3. Values associated with each link represent transfer rates or physical delay of the link. TCP and UDP sessions in the scenario are derived from pre-defined protocol set in NS–2. In this example, we have to modify the pre-defined RTP protocol so that it can control network congestion because the pre-defined RTP protocol cannot control it. According to TFRC3448[7], we have modified the behaviors of RTP senders and receivers as follows; the receiver can return appropriate feedback packets to
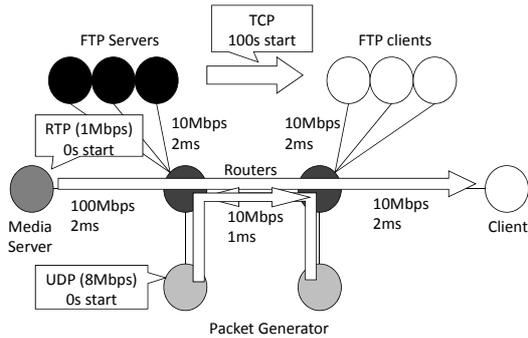
Figure 3: A Simulation Scenario Model of NS–2



Figure 4: Comparison of the Througput Obtained with NS-2 and PRISM



Figure 5: Comparison of Packet Loss Rates

the sender, which include information of a packet loss rate, network delay and so on, and the sender can control a transfer rate due to the feedback packets. In our implementation, we have added about 500 lines of code to eight files (twelve methods) of NS–2 sources.

## 6 Experiments

We have performed some experiments using our PRISM and NS–2 models described in Sec.4 and 5. The experiments were performed under an environment of Windows Vista 64 bit, Intel Core 2 Duo 2.33GHz, and 2.00GB of M.M.

In the experiments, we assume packet transmission parameters as follows: the packet size is 500 Byte, the number of packets acknowledged by a single TCP acknowledgment is one, and the TCP retransmission timeout value is $4 \times RTT$ second. Also we consider three scenarios with respect to the buffer size of routers. In the scenarios, the buffer sizes are 32, 64, and 128 KB respectively.

In the paper, we have performed two experiments. In the first experiment, we performed one trial run for PRISM and NS–2 simulation, respectively, and compared the simulation results. In the second experiment, we performed about 400 trial runs for PRISM simulation, and created a more simplified PRISM model based on the simulation results. Using the simplified model, we verify the minimum and maximum throughput of the RTP session. To make the simulation results more reliable, we chose about 400 thousands as the number of trial runs, which is default number trial runs for PRISM simulation.

### 6.1 Comparison of the Simulation Results

Here, we present the simulation results of the buffer size of 64KB in Fig.4 and 5. Throughput in the graph means the average throughput within one second, and a packet loss rate means a calculated value at the time as defined in [7]. In the scenario of the example, file transmission starts after 100 seconds from the start of RTP session, and this causes the network congestion. Consequently, the throughput of the RTP session goes down and the packet loss rate of it comes up. The simulation results of Fig.4 and Fig.5 show that our PRISM model and NS–2 model behave similarly even if the network congestion occurs.
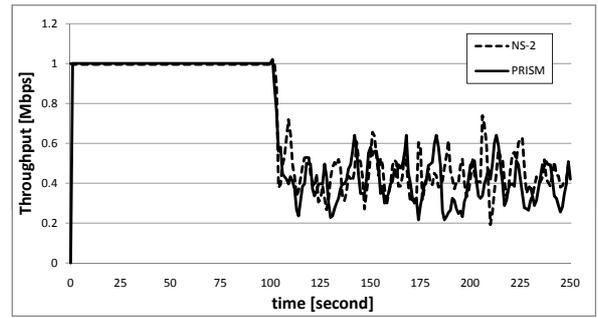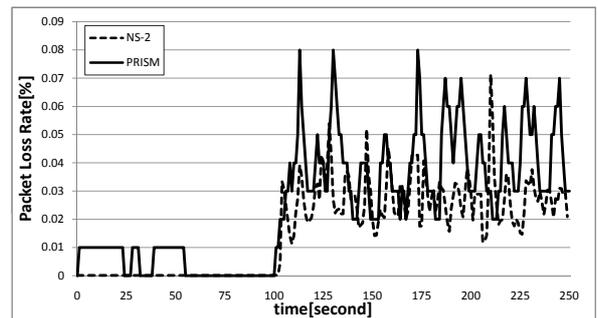
To analyze the correctness of our PRISM model in detail, we compare the average, variance, minimum and maximum of throughput measured by PRISM and NS–2. Table 2 represents the analyzed data. Though the maximum throughput of PRISM and NS–2 are similar in all scenarios, the average and minimum throughput are a little bit different in the scenarios of 32KB and 128KB. We think one of the reasons is that we abstract a packet sending mechanism in the PRISM model, that is, when the packet transmission rate is high, our model transmits a number of packets at a time. We think this may cause the differences of behaviors between the PRISM model and NS–2 one. However, Table 2 shows that the behaviors of these models seem to be totally similar.

For one trial run of the PRISM simulation, it takes about two or three seconds, while it takes about three minutes in the simulation of NS–2.

### 6.2 Verification results for the simplified model

Here, we verify minimum and maximum throughput that the media server can provide in the worst case and best case. In the verification, we use a simplified model based on the

Table 2: Comparison of the Analyzed Data

| Buffer size | 32KB | | 64KB | | 128KB | |
|---|---|---|---|---|---|---|
| Model | NS-2 | PRISM | NS-2 | PRISM | NS-2 | PRISM |
| Maximum | 607 | 588 | 738 | 640 | 995 | 980 |
| Minimum | 203 | 48 | 193 | 216 | 33 | 144 |
| Average | 242 | 372 | 443 | 401 | 473 | 367 |
| Variance | 11 | 13 | 10 | 11 | 12 | 12 |
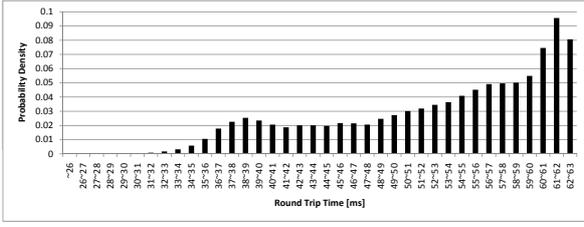
Figure 6: The Discrete Probability Distribution of Round Trip Time
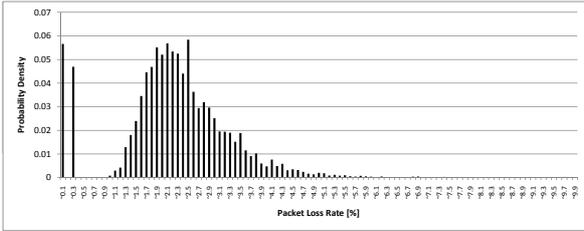


Figure 7: The Discrete Probability Distribution of Packet Loss Rates

simulation results on the detailed PRISM model. The simulation results should contain discrete probability distributions of the round trip time and the packet loss rates. To obtain reliable simulation results, we perform about 400 thousands of trial runs of PRISM simulation, which is default number of the PRISM simulation.

When the buffer size is 64KB, the discrete probability distributions of the round trip time and the packet loss rates are shown in Fig.6 and 7, respectively.

The verification property to be checked is given as follows; $P =?[Throughput < xUThroughput > x]$. This property means the probability of $Throughput$ being the value of $x$, where $Throughput$ is a variable to manage the throughput of the RTP session. The verification is performed with varying $x$ from 0 to 1000. Figure 8 represents occurrence probability of the throughput. The graph represents the probability that the throughput becomes greater than or equals to value of x-axis.

In the Fig.8, when the throughput is less than 116Kbps, the probability keeps the value one, and until the throughput becomes 1000Kbps the probability is greater than zero. From the verification result, it is guaranteed that the through-

put of the media server is greater than 116Kbps in the worst case. Also, the result shows that the media server can transmit packets with the rate of 1000Kbps in the best case. The simulation results of PRISM and NS–2 do not contradict to this verification result.
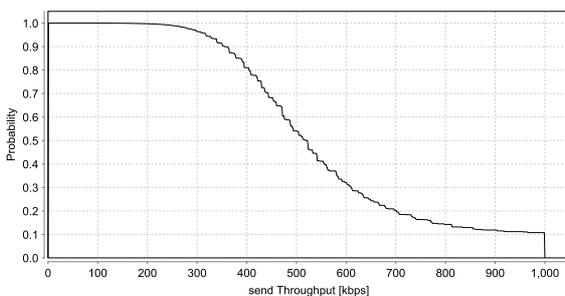
Although it takes about 10 hours to perform about 400 trial runs for PRISM simulation on the detailed model, in the verification on the simplified model, it takes about 40 seconds.

## 7 Discussion

In our case study, we have verified throughput of the media server using the detailed model and simplified model. The detailed model includes behaviors of the buffer control of the router and behaviors of the RTP, TCP and UDP protocols in detail. Although the detailed model is too complicated to verify qualitative properties of the system, we can measure some network properties using the simulation function of PRISM. On the other hand, the simplified model focuses on the verification of the minimum and maximum throughput, which is considered as one of important properties of the real-time systems. The simplified model is based on the throughput equation shown in Sec.2, where the dominant parameters such as a packet loss rate and round trip time are derived from the simulation results on the detailed model. The verification on the simplified model has been performed in a few seconds.

From the experiments, we can reconstruct our method as follows. The method will be useful to verify the minimum and maximum throughput that a RTSP connection can provide in an environment where some TCP and UDP connections exist simultaneously. In our method, first, we create a detailed model of the system which should have a buffer control mechanism of a router considered to be a bottleneck, as well as network congestion control mechanisms of RTSP and TCP. Next, using the simulation function of PRISM, we measure distributions of round trip time and a packet loss rate on the detailed model, and create discrete probability distributions of them. Based on the probability distributions and the throughput equation shown in Sec.2, we create a simplified model. The simplified model decides its round trip time and packet loss rate according to the probability distributions of them. Finally, the minimum and maximum throughput can be verified on the simplified model by probabilistic model checking with PRISM.

In our simplified model, we assume that packet loss rates and round trip time are decided independently. In fact, these parameters are not independent, i.e., when network congestion occurs, packet loss rates and round trip time tend to increase simultaneously. In the future work, we need to create a simplified model which represents their dependency.

Though simulation with NS–2 or PRISM can evaluate the system's performance, the simulation results depend on paths evaluated in the simulation. Therefore, from the simulation results, we cannot prove qualitative properties of the system. If we obtain a occurrence probability of throughput from the results of Fig.4, we cannot say it is reliable. In the paper, we created the simplified model and we can verify the qualitative properties of the system using model checking.



Figure 8: Occurrence Probability of Throughput

## 8 Conclusion

This paper presents a discrete Markov chain for a real-time distributed system, and evaluate its QoS properties using the simulation function of PRISM. To validate the correctness of our model, we also model it in a model for the well-known network simulator NS–2, and give the comparison of their simulation results. The comparison shows that the result of PRISM simulation is very similar to that of NS–2. It shows that the proposed method is useful to analyze the network performance. We think that such analysis is useful for other kind of network analysis.

The future works include automatic derivation of simplified model suitable for model checking analysis. Many abstraction techniques are proposed for model checking. We want to apply such techniques to the process.

## Acknowledgments

## REFERENCES

[1] A. Hinton and M. Kwiatkowska and 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.

[2] "The Network Simulator - ns–2," available at http://www.isi.edu/nsnam/ns/.

[3] E. M. Clarke, O. Grumberg, and D A. Peled: "Model Checking," MIT Press, 2000.

[4] E. J. Kim, K. H. Yum, and C. R. Das : "Calculation of Deadline Missing Probability in a QoS Capable Cluster Interconnect," IEEE International Symposium on Network Computing and Applications (NCA'01), pp.36, 2001.

[5] R. Rejaie, M. Handley, and D. Estrin : "RAP: An end-to-end ratebased congestion control mechanism for realtime streams in the internet," in Proc. of INFOCOM 1999, pp.1337–1345, 1999.

[6] I. Rhee, V. ozdemir, and Y. Yi : "TEAR: TCP emulation at recievers – flow control for multimedia streaming," NCSU Technical Report, 2000.

[7] M. Handly, S. Floyd, J. Padhye, and J. Widmer : "TCP friendly rate control (TFRC): protocol specification," Request for Comments (RFC)3448, 2003.

[8] M. Kwiatkowska, G. Norman, D. Parker and J. Sproston : "Performance Analysis of Probabilistic Timed Automata using Digital Clocks," In Proc. Formal Modeling and Analysis of Timed Systems (FORMATS'03), 2003.

[9] M. Kwiatkowska, G. Norman and J. Sproston : "Probabilistic Model Checking of the IEEE 802.11 Wireless Local Area Network Protocol," In Proc. PAPM/PROBMIV'02, volume 2399 of LNCS, pp.169–187, 2002.

[10] M. Kwiatkowska, G. Norman, J. Sproston and F. Wang : "Symbolic Model Checking for Probabilistic Timed Automata," In Proc. FORMATS/FTRTFT'04, volume 3253 of LNCS, pp.293–308, 2004.

[11] H. Hansson and B. Jonsson : "A logic for reasoning about time and probability," Formal Aspects of Computing, 6(5)pp.512–535, 1994.

[12] A. Aziz, K. Sanwal, V. Singhal, and R. Brayton : "Verifying continuous time Markov chains," LNCS 1026, pp.499–513, 1996.

[13] R. Alur, T. A. Henzinger and M. Y. Vardi : "Parametric Real-time Reasoning," Proc. 25th ACM Annual Symp. on the Theory of Computing (STOC'93), 592–601,1993

[14] Y. Taniguchi, A. Ueoka, N. Wakamiya, M. Murata, and F. Noda, "Implementation and Evaluation of Proxy Caching System for MPEG-4 Video Streaming with Quality Adjustment Mechanism," in Proc. of The 5th AEARU Workshop on Web Technology, pp.27–34, 2003.

[15] E. Nagai, K. Okano, S. Kusumoto: "Performance verification considering the network delay for Real-time Distributed Systems with the probabilistic model checker PRISM (in Japanese)," IEICE Technical Report, vol.105, no.597, pp.19–24, 2006.

[16] D. Mahrenholz and S. Ivanov: "Real-Time Network Emulation with ns-2," In 8th IEEE International Symposium on Distributed Simulation and Real-Time Applications, ser. 8. IEEE Computer Society, October 2004, pp. 29-36.

[17] M. Yajnik, S. Moon, J. Kurose, and D. Towsley: "Measurement and modeling of the temporal dependence in packet loss," In Proc. INFOCOMM'99, 1999.

# Network Mobility Management based on NEMO/MobileIPv6-MIB

†Naoki Nakamura, *Kazuhide Koide, ‡Takafumi Maruyama, ‡Debasish Chakraborty,
**Glenn Mansfield Keeni, ‡Takuo Suganuma, ‡Norio Shiratori

†School of Medicine, Tohoku University, Japan
‡Research School of Medicine, Tohoku University, Japan
* KDDI Corporation, Japan
** Cyber Solution Inc., Japan
{nakamura,maruyama,deba,suganuma,norio}@shiratori.riec.tohoku.ac.jp
ka-koide@kddi.com, glenn@cysol.co.jp

*Abstract* - For conventional network management it is implicitly assumed that the managed object will not change its location. So only the static information is available, which is used for simple reporting. The scenario is quite different with mobile devices, as devices are inherently dynamic. It causes difficulties for monitoring processes of management information due to roaming of devices or wireless signal errors. So it has become imperative to review the current approach of network management. To overcome this problem, we propose a new network management technologies based on standard network management protocol. Our focus in this paper is mainly on two items: (i) location awareness, and (ii) intelligent monitoring scheme. Here, we explained about MobileIPv6/NEMO-MIB, which is aware of mobile device's dynamically changing location. We have also discussed the monitoring scheme of network management infomation in detail.

*Keywords*: MobileIP, NEMO, IPv6, SNMP

## 1 Introduction

Location of managed object is implicitly assumed as static for traditional network. The location related information, if available is static and its use is limited to reporting purposes. However, with mobile devices these attributes are inherently dynamic. So, it is important for network management to understand the dynamics of location of devices for planning, administering and controlling a network. When planing, distribution or managing configuration of parameters in mobility protocols, density, roaming frequency and visiting duration of devices in the network are the important piece of information. So it is essential to review the current approach of management and further reconsideration is required for configuration, operations, performance and security management.

The basic concept of *reachability* has changed due to the introduction of mobility of devices. An IP network can be connectionless and the routes may change dynamically. But in conventional thinking, under normal circumstances network devices are considered to be reachable. The same assumption is extended to the manager-agent management framework as well. In monitoring process, polling at a regular interval is conducted to collect management information. This interval is generally set larger than the response time, so that the time stamp of the polling request or the response can be considered as the time-stamp of the information.

There is a wide fluctuation of response time in mobility aware networks due to the instability of reachability and this can cause severe degradation of quality of collected information. Though there are certain conveniences of network mobility, considering NEMO environment, but optimized route can not always be used. In that case, the amount of polling traffic could be substantial. With available bandwidth for wireless network is usually narrow, it can cause severe network congestion on certain links. In this paper, we propose a network management framework, which takes into consideration the inherent dynamic nature of location and the instability. The rest of the paper is structured as follows. In Section 2 we describe the problem statement. In Section 3 we presents our proposal on MobileIPv6/NEMO-MIB and two monitoring schemes. Evaluation is discussed in Section 4 and in Section 5 we conclude our work.

## 2 Problem Statement

### 2.1 Network Model

MobileIPv6[1] and NEMO (Network Mobility Support)[2] realize node and network mobility in IPv6 Internet.

The MobileIPv6 architecture is described in terms of three types of entities: mobile node (MN), correspondent node (CN) and home agent (HA). When a MN roams from one network to another, the IP address changes. This address is called Care-of Address (CoA). Each MN has its own IP address authorized by its home network, and called Home Address (HoA). When CoA of MN changes, MN registers it to HA with HoA. HA maintains the registered CoA/HoA sets of MNs. It is called *binding cache*. When packets from CNs destined to HoA come, HA forwards them to the registered CoA.

NEMO is a simple extension of MobileIPv6, adding an entity called mobile router (MR). A MR is basically a router with the additional functionality of mobility support that normal routers does not have. A MR registers its network prefix to the binding cache. The prefix is called Mobile Network Prefix (MNP). Nodes connected to MNP can have mobility with MR.

In this paper, we call MN and MR as *mobile devices*. Both MobileIPv6 and NEMO enable global mobility of mobile devices with no restriction of area.

## 2.2 Needs of New Management Framework

Generally management of network devices is carried out by monitoring or setting the value of a "Managed object(MO)". MOs are accessed by SNMP. The traditional SNMP-based network management adopts *manager-agent management framework*. Agents that employ MIBs are monitored by a manager or managers using SNMP. Fig. 1-(a) shows the overview of this framework. Fig. 1-(b) describes two additionally needed extensions, to take into consideration the inherent dynamic nature of location and the instable nature of network reachability.

First, for all practical purposes, the scope of the concept of location in traditional network management has been limited to the "sysLocation" in the MIB-II [3] that is the MO provided static descriptive value. The traditional MIB-II is insufficient for managing dynamic nature of location because the location of a device changes dynamically in networks that support mobility. So the new MIB should be aware of inherent dynamic nature of location of devices. Thus *location-aware* MIBs will be needed. Main concern is the concept of location of the mobile device, and how to monitor and track the device locations monitored continuously especially in MobileIPv6/NEMO environment. In section 2.3 we consider about how to monitor the location of the mobile devices.

The second need is to overcome the instable nature of network reachability. Monitoring process is crucial for many types of management information, because the change in the values of the MOs with time is of key interest. Monitoring performance of mobility protocols, mobile devices and applications on the devices is an important aspect of network management. Location information also should be monitored. It is important to carry out periodical and continuous polling precisely. But the wide fluctuation of the response time between the manager and the agents caused by roaming of the device is unavoidable. The transient failure of wireless links will cause many packet losses. SNMP polling suffers from lots of polling timeout and they degrade the quality of collected management information. It is difficult to maintain a simplified monitoring process. Monitoring process should be more *intelligent* to overcome instability of the reachability. In section 2.4 we consider about this problem and propose a novel technology for continuous device monitoring over an instable network.

The new dimensions, *location-awareness*, and *intelligent monitoring scheme*, added to networking by mobility require extensions to the traditional network management framework.

## 2.3 Location-awareness

We classify the concept of location of a mobile device based on these three aspects:

Organizational Location: the administrative domain to which the mobile device belongs.

Segment Location: the subnetwork to which the mobile device is attached.

Geographical Location: the longitude, latitude, altitude information of the mobile device.
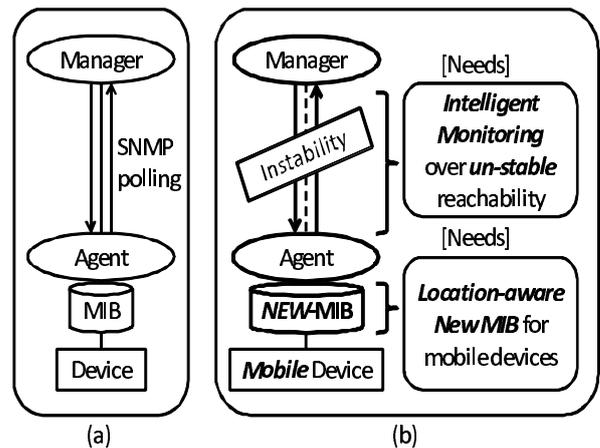


Figure 1: Manager-Agent management framework using SNMP (a)Traditional management: Manager monitors device's MIB. (b) Extensions needed: 1) new location aware MIB of mobile device and 2)intelligent monitoring scheme that realize monitoring over instable reachability.



| Location of Device | Organization | Segment | Geographic |
|---|---|---|---|
| Type-(1) | Unchange (A->A) | *Change(a->b)* | *change* |
| Type-(2) | *Change(A->B)* | Unchange(b->b) | *change* |
| Type-(3) | *Change(B->C)* | *Change(b->c)* | *change* |
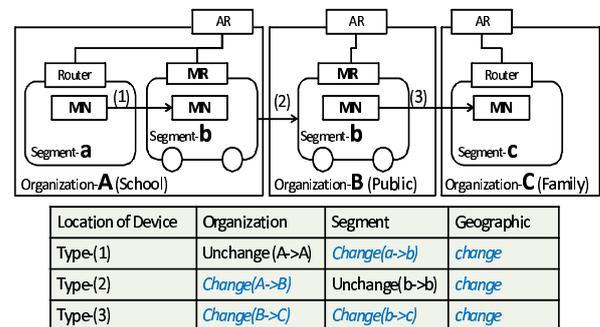
Figure 2: Three types of location changes in a mobile node: Organization/Segment/Geographical location changes occur separately.

In the traditional network, all aspects of location of a device are fixed. Fig. 2 describes three separate scenarios of location changes for mobile devices in MobileIPv6/NEMO network environment. We can see that device and network mobility allows three different location aspects to change independently.

In scenario type-(1) the MN moves from a segment under a router (inside the school) to the other segment which is under the MR (mobile router deployed inside the school-bus) within the same organization (school network). In type-(2) the MN itself does not move, the MR leaves from one organization (school network) and connects to the other organization (public network). In type-(3) the MN moves from the segment (school-bus network that attaches to the public network) to another segment that belongs to the different organization (family network).

As described above, these location changes occur dynamically and independently, sometimes without user's intention. Looking current locations of mobile devices enables managers to get the current view of their managing devices, and the network.

## 2.4 Intelligent monitoring scheme

It is important to carry out periodical and continuous polling precisely in monitoring process. However, in general the concept of "time" in traditional monitoring does not have an exact definition.

As Fig. 3-(a) shows, in some cases it is the time when the request was sent by a manager. In others it is the time when the response is received. It is considered to be an unnecessary overhead to tag every observation with the actual time of the observation. In traditional networks the RTT between the manager and the monitored agent is generally of the order of 10ms. On WLANs it can goes upto the order of 100ms. The polling interval $t$ is of the order of several seconds. So, the inaccuracy introduced in the time-stamp may be considered to be negligible. But in a mobility-aware network, the RTT may vary widely. Typical latency of hand-over within same wireless access media is 300-500ms (L2 movement), and at least 3 seconds (L3 movement) [4] without any optimization. Thus information collection by polling at *regular intervals* in the traditional mode may cause an accuracy degradation of periodical information. Additionally, a traditional SNMP manager uses RTT information to fix the polling timeout. In the case of periodical polling, this approach results in a large number of timeouts and consequent data loss in data collection. This causes severe degradation in the quality of collected information.

To overcome these problems, we already proposed the *timestamped monitoring technique*. Fig. 3-(b) shows the brief structure of this technique. The *bulk retrieval technique* can be realized with *timestamped monitoring & buffering* proposed in [5] to solve the problem of data loss. Data buffering at the agent-side recovers data losses in polling.

In NEMO environment, where route optimization is not supported, all packets toward MN are forwarded through HA because packets from Manager to MN are encapsulated in bi-directional tunnel between HA and MR. Therefore transferring the excess of management information may causes performance degradation of HA network. So, if possible, it would be better to monitor MN without going through this tunnel, i.e connect directly. When MobileIPv6 supports route optimization, manager just monitor the MN.

Fig. 4 explains another problem intuitively. During MR's long term disconnection, there will be large amount of data buffered at agents. Just after reconnection, agents may send them all together and there will be a sudden surge in monitoring traffic. It causes congestion of MR's upstream link and may seriously affect to other communications of MNs. If MR moves into a narrow bandwidth environment, the affect of this problem become severe. We need to control the timing or the amount of monitoring traffic to avoid this.

## 3 Proposal

### 3.1 MobileIPv6/NEMO-MIB

To manage location, we proposed *MobileIPv6-MIB* [6] and *NEMO-MIB* [7] objects in the MobileIPv6/NEMO-MIB as shown in Table 1.
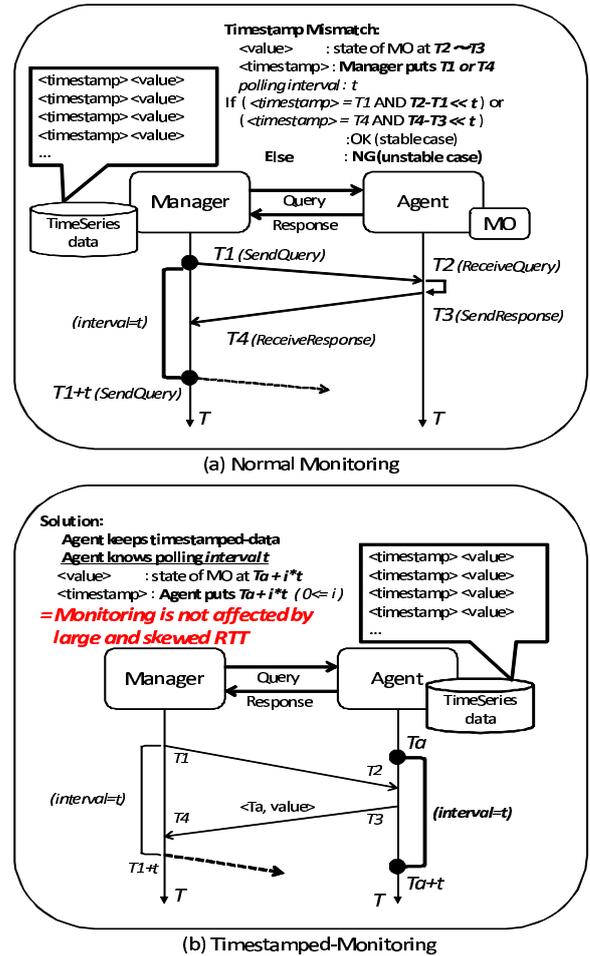


Figure 3: Timestamped monitoring - (a) Normal periodical monitoring contains risk of time-stamp mismatch (b) Agent itself puts time-stamp

A mobile device's global address (CoA) potentially reveals address of the network to which the device is attached. Thereby the location of the mobile device (the organization and/or the segment that the device exists in) can be traced. Location information can be collected from *binding information* available at the HA that is communicating with the mobile device (MR/MN).

The MobileIPv6/NEMO-MIB can provide location information of mobile devices, by using MNPs of MRs and MR/MN's CoAs.

MobileIPv6-MIB is the MIB module for MobileIPv6 entities. *mip6BindingCacheTable* reveals the binding cache information of MNs that is maintained in the HA (or in the CN). This reveals the segment that the MN is connecting. If the prefix of the CoA is from static network, the organization where the MN is connecting is also revealed. *mip6BindingHistoryTable* has the expired binding cache information. We can track the movement of mobile devices from the history. *mip6MnBLTable* is the list of binding update that is maintained by the mobile device itself. It shows the attempts of binding update by the MN. It is possible by using this table to get information of CoAs even after some binding updates had been failed.

*NEMO-MIB* is the MIB module for NEMO entities. *nemoBind-*
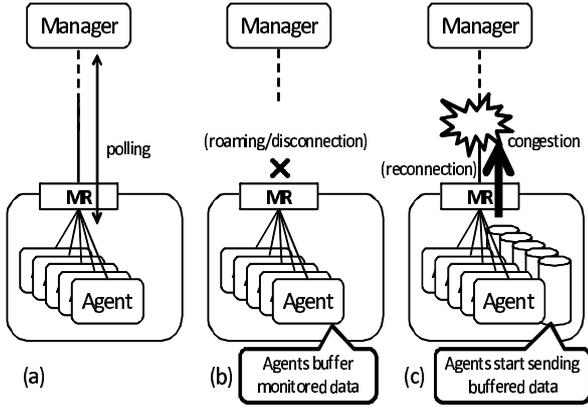
Figure 4: (a) Normal polling (b) Agents buffer data during MR's disconnection (c) Data sending after reconnection causes a congestion of MR's upstream
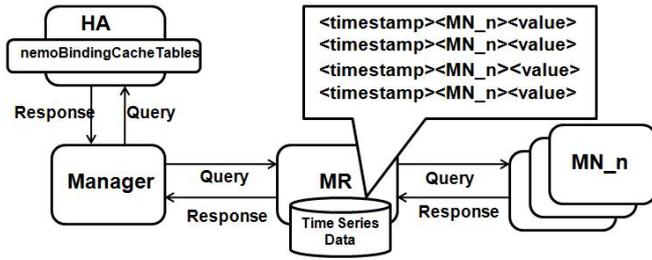


Figure 5: Overview of direct monitor method

*ingCacheTable* is extended with *mip6BindingCacheTable*. It will serve the segment/organization information of MRs. NEMO-MIB also defines the *nemoHaMobileNetworkPrefixTable* that contains information of the registered MNP in the MR's network. If the mobile device has the CoA from the MNP, it can be revealed that the mobile device is connecting to the MR's segment.

These give rise to a new area of management - *location management*, wherein the location of a mobile device is monitored, analyzed, utilized and controlled, as new applications are emerging in this area. It helps to diagnose the cause of the faults, performance degradations, and so forth. Managers have to find and monitor changes in these locations all the time.

## 3.2 Direct Monitoring

To collect management information effectively, we propose a direct monitoring method which store the managed information on MR for MNs.

In NEMO environment which does not support route optimization, packets from Manager to MN are encapsulated in bi-directional tunnel between HA and MR. As all the packets toward MN are forwarded through HA, manager can not connect MN directly. As manager can connect CoA of MR directly, in our proposed method we are considering about fetching the management information from MN through MR.

Fig. 5 illustrates the overview of the direct monitoring method. At MR we use buffering scheme like mobisnmp[5]. Its managed values are buffered not only time-stamp but also identifications of each MNs. Here MR periodically monitors the managed value from MN and store it in its buffer. When manager tries to request the buffered managed information, first, the manager picks up a MR's CoA in the nemoBindingCacheTables of HA's NEMO-MIB. Then it collects managed information buffered in MR. Since MR's CoA may change frequently, manager always makes continuous efforts to pick it up. This method can omit wasted managed traffic, and thus can lead to higher performance, especially when bandwidth of HA is narrow.

## 3.3 Delayed Data Sending Method

We propose a novel method to prevent congestion on MR's upstream link caused by SNMP agents that buffered large amount of management information and send them at a burst. We call this method as *delayed data sending method*. The idea is illustrated in Fig. 6. This method is based on direct monitoring method.

The MR monitors the polling interval that managers requests for managed information. When the MR's upstream link is disconnected, it also monitors the duration of its disconnection. The MR has the configured value of $L$. $L$ is the permitted ratio of MR's upstream bandwidth $U$ for monitoring. Monitoring traffic can consume at most $U * L$. This value is configured to avoid congestion of the upstream link. The MR estimates the amount of monitoring traffics to be transmitted after the upstream is reconnected. If monitoring traffic $M$ assumes to exceed the limitation $U \times L$ during next transmission, proposed method decides how much transmission should be delayed and how many times information data should be divided so that monitoring traffic would not exceed the limit.

Considering the interval of monitoring traffic, amount of its traffic, and disconnection duration, MR estimates the amount of monitoring traffic traveling through MR at each time. If total amount of monitoring traffic $M$ assumes to exceed the limitation $U \times L$, MR delays and divides the transmission into blocks with following procedure. In our scheme, we take an approach so that each traffic will be distributed fairly.

If monitoring traffic exceeds the limitation $U \times L$, MR firstly divides the information data into $\lceil M/U \times L \rceil$ blocks, for example, $B_1$, $B_2$, $B_3$. Then transmission time of divided blocks are arranged so that $B_2$, $B_3$ will be transmitted at $t+u$, $t+2u$, where $t$ is the transmission time of $B_1$ and $u$ is the unit of time. If maximum amount of arranged monitoring traffic $M$ still exceeds the limitation $U \times L$, then each block of transmission is divided into two. Then transmission time of divided blocks are re-arranged again. In spite of the adjustment, if it exceeds the limitation, MR plans to divide block of transmission into three. In this way, re-arrangement of the delay and division of the transmission will be done repeatedly until maximum amount of monitoring traffic $M$ is less than the limitation $U \times L$.

Fig. 6 shows a very simple case to explain it. There are 4 nodes, a, b, c and d. In this case, for node c and d, MR

Table 1: Managed Objects in MobileIPv6/NEMO-MIB related to location management

| | *Name of the object* | *Entities to maintain* | *Description* |
|---|---|---|---|
| MobileIPv6-MIB | mip6BindingCacheTable | HA/CN | Models the Binding Cache |
| | mip6BindingHistoryTable | HA/CN | Tracks the history of the Binding Cache |
| | mip6MnHomeAddressTable | MN(MR) | List of Home Addresses pertaining to the MN(MR) |
| | mip6MnBLTable | MN(MR) | Models the Binding Update List |
| NEMO-MIB | nemoBindingCacheTable | HA | Extended Binding Cache in NEMO |
| | nemoMrBLTable | MR | Extended Binding Update List in NEMO |
| | nemoHaMobileNetworkPrefixTable | HA | List of Mobile Network Prefix registered by HA |
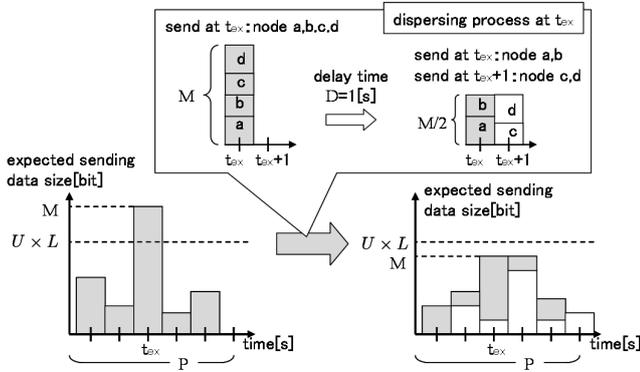


Figure 6: Delayed data sending method



Figure 7: Topology of the experimental network: It includes 2 foreign networks: the native-IPv6 network and the cellular network which is non-native-IPv6 network.

waits for 1 sec to send the data. As a result, $M$ has been successfully controlled not to exceed the limit ($U * L$). Delayed data sending method controls transmission for large amount of buffered management data. If MR moves into a narrow bandwidth environment, our method becomes more effective. A more detail description of our method can be found in [8].

## 4 Evaluation

### 4.1 Experimental Environment

We build an experimental MobileIPv6/NEMO network in our laboratory, using SHISA [9], one of most widely-used MobileIPv6/NEMO implementations. Fig. 7 shows the structure of the network. IPv6 is deployed in the entire network. One single MR is attached with a few MNs. It can moves from one network to another. Each network has an Access Router (AR).

There are 3 rooms, each room has different types of access network. In Room-1 the AR connects to the home network, that is, where the HA of MR exists. Its structure is almost same as a simple LAN. The IPv6 upstream of our laboratory is *JGN2-IPv6*. In Room-2, it is a foreign network, and the upstream is via cellular network. Here IPv6 connection is non-native, that is, the IPv6 connection is provided by tunnel technology. Also Room-3 is a foreign network but here it natively connects to IPv6. It connects to the *SINET-IPv6*. In Room-1 and Room-3, ARs are wire-connected. But in Room-2, the AR has a cellular network interface and is connected with the wireless link.

These networks can be divided according to its performance. In Room-1 RTT and packet loss rate between HA and MR is low. In Room-2 the RTT between HA and MR is big and there

can be high packet loss. In Room-3 packet loss rate between HA and MR is low, but RTT is medium.

### 4.2 Binding related data collection using prototype system

Fig. 8 shows the result of polling by the prototype SNMP manager from the prototype MobileIPv6/NEMO-MIB module implemented in SNMP agents running on HA and MR. Fig. 8-(a) shows *mip6BindingCacheTable* (binding cache table) from the HA. Fig. 8-(b) reveals *mip6MnBLTable* (binding update list table) collected from the MR that corresponds to (a).

We confirmed that the manager can get correct information about bindings from MIB modules in every case that the MR/MN connects to one of the 3 rooms in our experimental network. This result reveals that our MIB implementation can provide correct binding related information to managers without serious trouble, in a reasonable MobileIPv6/NEMO environment. To our knowledge, this is one of the pioneering work on implementation of MobileIPv6/NEMO-MIB. It demonstrates the feasibility of monitoring mobile devices using the Internet standard network management protocol, SNMP. Apart from the traditional traffic or configuration information, the MobileIPv6/NEMO-MIB does contain information

(a) snmpwalk to the HA: Binding Cache information

(b) snmpwalk to the MR: Binding Update List information
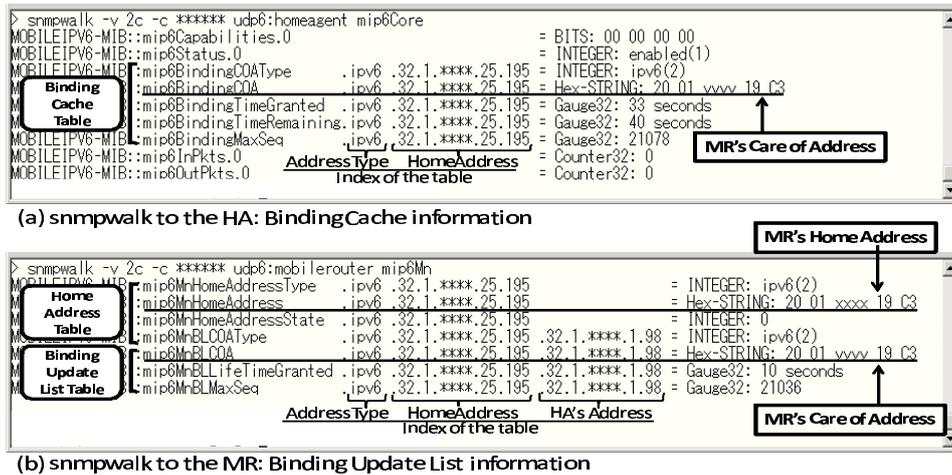
Figure 8: Screenshot of SNMP polling of the implemented prototypes

that provides location information in one form or the other for monitoring mobility. Information that can be collected from MobileIPv6/NEMO-MIB is mandatory in developing new management applications for location-aware network management in MobileIPv6/NEMO network. The example of application is explained in section 4.4.

## 4.3 Delayed data sending method

To verify the effectiveness of our proposed *delayed data sending method*, we compared the performance by using ns-2 [10].
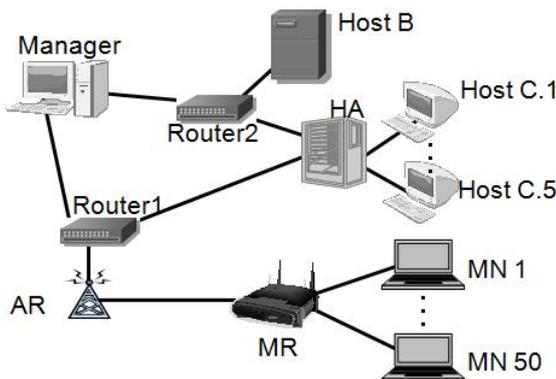


Figure 9: Simulation environment

Simulation environment is shown in Fig. 9. There are 50 MNs connected under the MR. We configured that, 1) manager polls MOs in 30 sec interval, 2) each MN collects 10 MO per 5 sec, 3) the size of each MO is 100 bytes, 4) five hosts (C1, C2, C3, C4 and C5) are connected at HA to produce background traffic toward host B, 5) MR's bandwidth between AR and MR is configured as 11 Mbps, and others are set at 50 Mbps. Simulations time is 600 seconds. To increase the amount of monitoring traffic bursty, link between MR and AR goes down from 90 to 210 sec.

In the above described environment, two kinds of simulation scenarios are examined. In scenario (1), VoIP traffic is

transmitted from one of the MN to Host B at 1Mbps. In scenario (2), five hosts (C1, C2, C3, C4 and C5), which are connected under HA, transmit traffic to B at 10Mbps respectively. The limiting value $L$, the permitted ratio of MR's upstream bandwidth $U$ for monitoring, is configured as $0.1 (= 10\%)$.

By scenario (1), we confirm that proposed method can keep the monitoring traffic to a constant level and lead to the reduction of interference on VoIP traffic in terms of throughput and jitter.
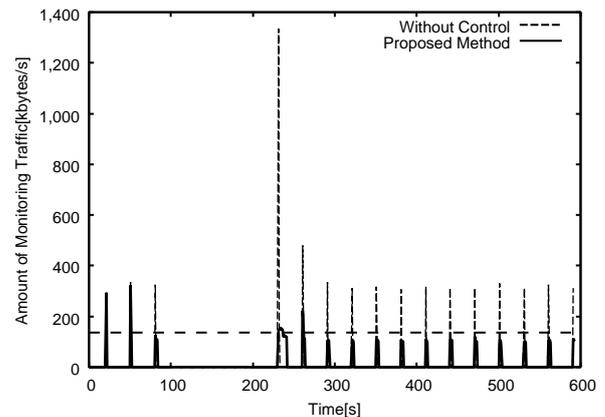


Figure 10: Monitoring traffic from 0 to 600 sec

Fig. 10 shows monitoring traffic during simulation. Traffic details just after the re-connection is shown in Fig. 11. These graphs show that, without controlling, the amount of traffic violates the limitation. On the other hand, our proposed method hardly exceeds the limit. Therefore our proposed method can effectively distribute bursty management data transmission for avoiding congestion on upstream link of MRs.

To confirm that proposed method can reduce the interference on VoIP between MN and host B, throughput and jitter of VoIP are measured as shown in Fig. 12 and 13. These results show, that by suppressing the amount of monitoring traffic to a lower level, proposed method can enable VoIP not only to achieve higher throughput but also can keep jitter low.
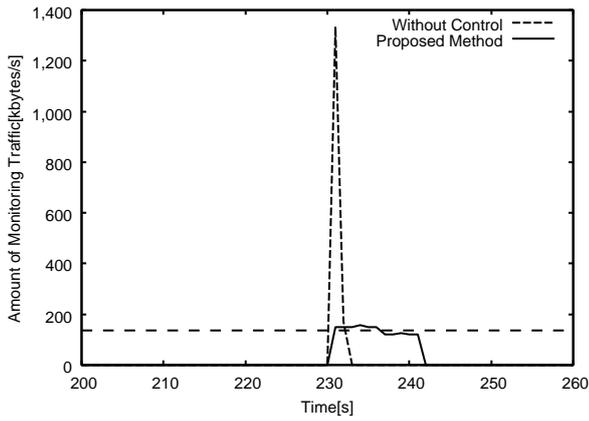
Figure 11: Detail of monitoring traffic right after link between MR and AR has recovered
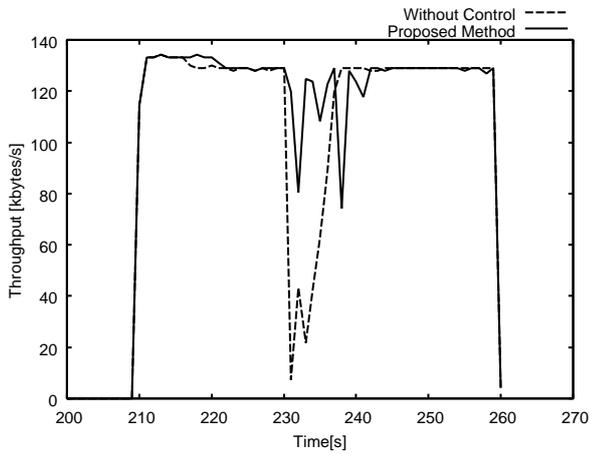


Figure 12: Comparison of throughput when the transmission is between MN to host B
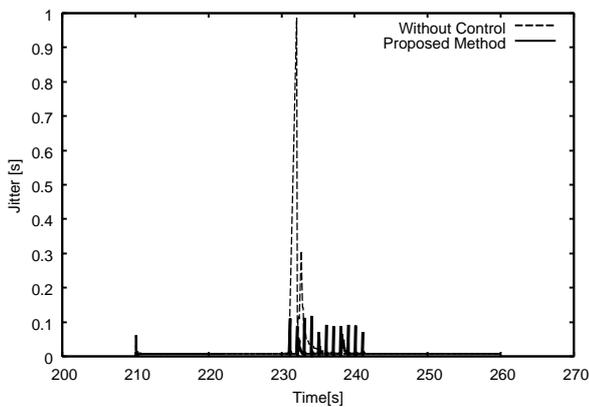


Figure 13: Comparison of jitter when the transmission is between MN to host B

Therefore, the proposed method can reduce the interference on other traffic under MR during monitoring.

Finally, to show the effect of direct monitoring, we measure the traffic from one host connected under HA to host B. Traffic on host C3 and host C4 is shown in Fig. 14 and 15 respectively. Both results show that proposed method is free from interference of monitoring traffic, and thus achieved higher throughput. This means proposed scheme is capable of reducing the interference on HA traffic effectively.
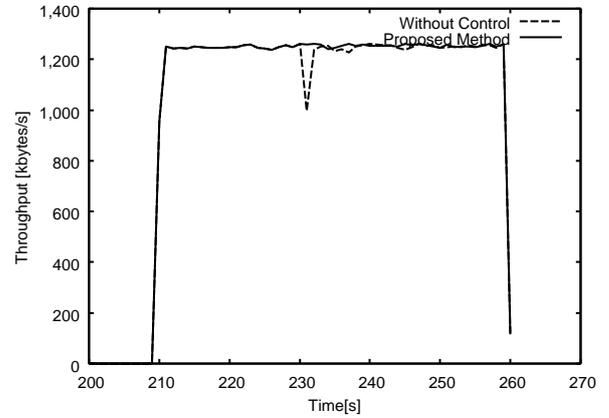


Figure 14: Comparison of throughput when transmission is between C3 to host B
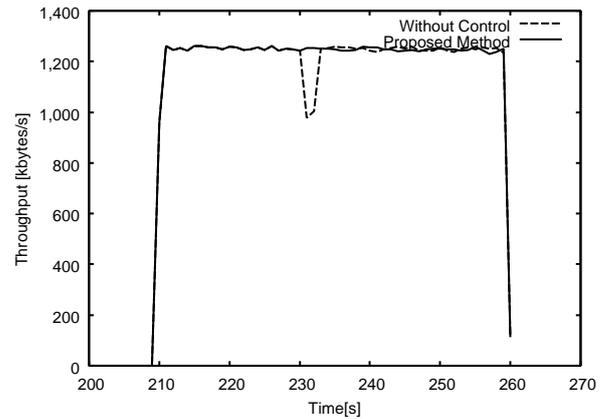


Figure 15: Comparison of throughput when transmission is between C4 to host B

## 4.4 Effective management application using location

As explained in this paper, the location-awareness and the intelligent monitoring scheme are novel and important axes of network management technologies.

In this section we explain about the example application now we are constructing.

The application will be apart from the traditional simple demonstration of network traffic and configuration information. The application uses location information to track the path of a mobile device. The MobileIPv6/NEMO-MIB modules can provide information of the network that the mobile
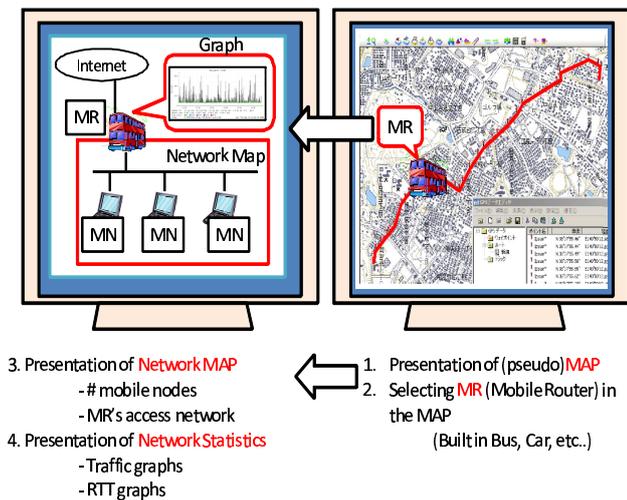
Figure 16: The screenshot of the application that realizes location-aware network management

device is currently connecting. It also provides the list of networks that the mobile device had been connected in the past. Three types of location of the mobile device can be determined with reasonable accuracy using other MIBs that service geographical location information like GPS-MIB. This opens up a new vista of tracking mobile devices for management. The application also shows how the path of a mobile device can be tracked even when it is out of network reach. It demonstrates that mobility-aware networks can be managed by using already standardized and widely utilized management protocol same as any other device and network. The mobility related status, parameters and metrics can be evaluated. The traffic can be monitored and the network configuration can be discovered.

We envisage the application that supports dynamic network topology visualization. Fig. 16 is a screenshot of the application. In this image, a MR is deployed on a public transport, like bus. GPS equipment is connected to the MR and it makes possible to locate the bus on the geographical map. This works only as an interface to point out for human managers that manages MRs. When a human manager points the bus, the application shows the information of the MR and its network. By using the information from MobileIPv6/NEMO-MIB, it is possible to show the network the MR is attached now. The information of MNP shows prefixes that is reachable under the MR. It is also possible to show the number of MNs connected to the MR. These functions reveal the network topology of MR's network. Traffic information of MR's network is also possible to provide.

## 5 Conclusion

For a new network management technologies of MobileIPv6/ NEMO network environment, we propose a *location aware* and intelligent monitoring scheme. We explained about the *MobileIPv6/NEMO-MIB* which is aware of mobile device's inherent dynamic nature of location, and we introduce an intelligent monitoring scheme. This is one of the first applications for monitoring mobile devices and networks that are completely based on Internet Standard network management protocols of MobileIPv6/NEMO. It expands the scope of network monitoring to mobile devices and networks, and also demonstrates new areas of management e.g. *location management*, in which network monitoring can be applied.

## REFERENCES

[1] D. Johnson, C. Perkins, and J. Arkko, "Mobility support in IPv6," Jun. 2004, RFC 3775.

[2] V.Devarapalli, R.Wakikawa, A.Petrescu, and P.Thubert, "Network mobility (NEMO) basic support protocol," Jan 2005, RFC 3963.

[3] K. McCloghrie and M. Rose, "Management information base for network management of tcp/ip-based internets: MIB-II," Mar. 1991, RFC 1213.

[4] C.T.Chou and K.G.Shin, "Smooth handoff with enhanced packet buffering-and-forwarding in wireless/mobile networks," *Wireless Networks*, vol. 13, no. 3, pp. 285–297, 2007.

[5] K.Koide, G.Kitagata, H.Kamiyama, D.Chakraborty, G.M.Keeni, and N.Shiratori, "MobiSNMP - a model for remote information collection from moving entities using SNMP over MobileIPv6," *IEICE Transactions on Communications*, vol. E88-B, no. 12, pp. 4481–4489, 2005.

[6] G.M.Keeni, K.Koide, K.Nagami, and S.Gundavelli, "Mobile IPv6 management information base," Apr. 2006, RFC 4295.

[7] S.Gundavelli, G. M.Keeni, K.Koide, and K.Nagami, "Network mobility (NEMO) management information base," Apr. 2009, RFC 5488.

[8] T.Maruyama, K.Koide, G.M.Keeni, and N.Shiratori, "An efficient method of information collection for management in NEMO environment (in Japanese)," in *IEICE Technical Report MoMuc2008-43(2008-9)*, Sep. 2008, pp. 1–6.

[9] "SHISA," http://www.mobileip.jp/.

[10] "The network simulator 2 (ns-2)," http://www.isi.edu/nsnam/ns/.

# Development of a wearable communication recorder triggered by voice for opportunistic communication

Tomoo Inoue[*] and Yuriko Kourai[*]

[*]Graduate School of Library, Information, and Media Studies, University of Tsukuba, Japan
inoue@slis.tsukuba.ac.jp, s0512184@u.tsukuba.ac.jp

*Abstract* – Everyday communication is not always planned in advance. There is type of communication called "opportunistic communication" that happens unintended in any place and in any moment. Conventional video recording equipment cannot capture this type of communication because it has to be set up in advance. This paper presents development of a wearable communication recorder (WCR) for opportunistic communication. To a problematic issue of dealing with large amount of video data in life log systems, the proposed wearable recorder only records interpersonal communication. It gets recording cue by utterance of a user and records communication as far back as some 30 seconds of the utterance. This can reduce useless recording and can record whole communication session at the same time. From the analysis of interpersonal communication, 10 seconds of backward recording is suggested to be acceptable.

*Keywords*: Wearable computing, Communication Research, Opportunistic Communication, Nonverbal, Multimedia.

## 1 INTRODUCTION

Research on interpersonal communication has been actively conducted since 1950's, and its major progress was brought by video recording technology. Observation was almost the only way for a researcher to study interpersonal communication that never happens again before the use of video recording technology. Video recording technology enabled a researcher to investigate a piece of communication activity repetitively and to analyze it in fine time sequence by frame-by-frame observation. In other words, detailed communication analysis cannot be achieved without video recorded communication.

It is well known that there are two types of interpersonal communication. One occurs at predetermined time and place while the other occurs opportunistically without prior appointment. Interpersonal communication that has been analyzed so far by video data falls into the former. Conventional video recording cannot capture the latter opportunistic communication because video recording equipment should be set up before communication starts.

However, video recording equipment has progressed. Wearable video recording equipment can be found recently partly owing to wearable computing. It has been used in life log research, for example.

We propose a wearable communication recorder to get video data of opportunistic communication from the background mentioned above. Opportunistic communication can be recorded by using such wearable video equipment. Problem is how to deal with huge amount of video records when video is always recorded. This is a major research issue in life log. Our approach is to record opportunistic communication alone automatically.

The rest of the chapters are composed as follows. Related research is explained in chapter 2. Our recorder is proposed in chapter 3. Implementation of the recorder is described in chapter 4. Daily informal communication is examined for the design of the recorder in chapter 5. Initial investigation of opportunistic communication is reported in chapter 6, and basic performance of the recorder is validated in chapter 7. Conclusion is given in chapter 8.

## 2 RELATED RESEARCH

### 2.1 Opportunistic Communication

Opportunistic communication is a type of informal communication. Informal communication is known to be important to keep and nurture human relationships and to facilitate group work. Because of this importance, there are many systems to support informal communication. VideoWindow is an early research system to support informal communication between distributed office rooms by audio and video link [1]. Cruiser is an early research system to support informal communication between distributed desktops [2].

There are many other systems to support informal communication since then. However they are not for recording and analysis of opportunistic communication.

### 2.2 Communication Recording

Video recordings of communication for communication analysis have been made frequently.

Examples of such video use are analysis of multi-party conversation by video recordings [3] and analysis of body movement synchrony in psychotherapeutic counseling by video [4]. Video is recorded manually in these cases. Video recording of multi-party conversation in a meeting with automatic speaker identification has been realized recently

[5]. However, all these are for communication that is scheduled in advance and takes place at fixed location.

There are recording of communication in a limited area. The Active Badge system is an early system of personal location. A user wears an infra-red transmitter named Active Badge. Networked sensor that is installed one or more in each room detects the Active Badge thus tells the location [6]. From the closeness of two or more badges, a system user is able to know the meeting of people. The Bat system is subsequent of Active Badge system. Ultrasound is emitted from a transmitter that a user wears and a number of receivers are installed on the ceiling, which makes high precision of locating possible [7]. Automatic recording of interactions is done using infra-red markers on objects in an area and infra-red receivers worn on users [8]. All these can record interactions more or less, but markers are necessary. This means that recording area is limited.

Because opportunistic communication occurs at any place, we cannot apply these systems to record opportunistic communication.

### 2.3 Life Log Research

Life log system is an always-on recorder of various events, user's behavior, and operated objects. It gains very large data because of its always-on feature. Dealing with very large data is not easy. Additionally, most of the data is useless. Information retrieval from this very large data is a major topic of research because of this feature. Because of this, it may not be the best to record everything when the type of information to be needed is known. Our research is different from generic life log research and is focused on recording of interpersonal communication.

Mixture of video and sensor information has been also researched because image processing alone is not assured to recognize objects in the video or to understand context. Sensor information can be annotation to the video recordings. It has not focused on interpersonal communication.

## 3 PROPOSAL OF WEARABLE COMMUNICATION RECORDER

We propose a wearable communication recorder (WCR) to get video data of opportunistic communication. Because opportunistic communication does not happen at a fixed place, conventional video recording equipment cannot fulfill the requirement of mobility. People also do not know when opportunistic communication happens because it happens without appointment. Typical opportunistic communication is short talk when 2 people come across in the hallway. Accordingly, recording equipment should record interpersonal communication without limitation of time and place. To fulfill this requirement, wearable equipment is applied.

Recording everything like a life log system may cause problem of dealing with large amount of video data although it is easy to do so. Recording communication alone is an option to avoid this problem. However it is disturbing for a user to turn on and off of recording manually every time

communication happens. It is desirable that the system automatically turn on and off.

One of the important issues for the recording of interpersonal communication is detection of initiation of communication and completion of communication. For automatic recording of communication, utterance can be a cue. It is not so simple to start recording when detecting user's voice and stop the recording when not detecting the voice. It is known that communication usually starts before first utterance by eye contact or salutation [9]. In accordance with this, the recorder should record nonverbal cues that occur before verbal utterance. Detecting these nonverbal cues robustly is not very easy in itself, and it needs measuring instruments that are not easy to be wearable.

To this issue, our recorder applies mechanism of a driving recorder. A driving recorder is a device commonly used in business cars such as a taxi. It records video around the car as evidence when it detects impact by car accident or sudden braking. Because it is not good for evidence video to start recording from the impact, video data is buffered about a minute. The buffered data is saved on impact.

Recording of interpersonal communication is thought to be possible by voice detection and mechanism similar to the recording mechanism of a driving recorder. Voice detection can be the trigger of saving buffered video data that include nonverbal cues that come before verbal utterance. For stop recording, it is also thought to be possible by extending the recording some more time after not detecting voice.

With this explained mechanism, WCR is presumably able to record whole communication session in any time and place while reducing useless data.

## 4 IMPLEMENTATION OF WEARABLE COMMUNICATION RECORDER

### 4.1 Equipment

A USB camera for video communication (Microsoft LifeCam VX-6000) is used for the video recording. It has a 1.3 million pixel CMOS sensor that gives clear image. The view angle is 71 degree and wide enough to capture communication target. It is with automatic adjustment of white balance that responds to brightness. It is with directional noise-canceling microphone and good enough to record conversation.

Built-in directional microphone of the USB camera is used for recording communication. Another throat microphone is used for assuring detection of user's utterance.

These are connected to a small laptop PC (Fujitsu LOOX P70-XN), which records the input information.

Figure 1 shows appearance of a user when wearing these equipments.

### 4.2 Software

OS of the PC is Windows XP Professional. Program is written in C++. OpenCV 1.0 is used for video processing. Windows Multimedia API is used for audio processing.

Figure 1: Appearance when wearing the system.

WCR keeps detecting audio from the throat microphone. Video from the USB camera and audio from the built-in microphone of the camera are buffered for some 30 seconds. When the audio from the throat microphone becomes bigger than a certain threshold, it is recognized as utterance of the user of the system. Then buffered video and audio is saved to a file, which is supposed to be recording of nonverbal communication before the utterance. On the other hand, when the audio from the throat microphone becomes and keeps less than a certain threshold after recognizing utterance, it is recognized as completion of the utterance. Recording stops after a certain period of time then.

The video is encoded in real time and saved as MotionJPEG with resolution of QVGA. The each audio is saved as 8 bit monaural RIFF with 22kHz sampling rate.

# 5 INVESTIGATION OF DAILY INFORMAL COMMUNICATION

## 5.1 Aim

To know the appropriate buffering time in WCR, communication was investigated. Although the actual target of WCR is opportunistic communication, daily informal communication was investigated because the targeted communication cannot be captured without our system and because daily informal communication is thought to be more similar to opportunistic communication than other type of communication.



Figure 2: Devices used for recording daily informal communication: Omnidirectional camera (left) and conference microphone (right) Appearance when wearing the system.

## 5.2 Method

An omnidirectional camera (VS-C42U-300-TK) and a conference microphone (BUFFALO BSKP-CU201) (Figure 2) were set up at a room which is approximately 9 m x 6 m (Figure 3). Communication in the room was recorded 27.5 hours in 5 days. Video was recorded in 2048 x 1536 (QXGA) with 6 fps. Horizontal view angle of the camera is 10 – 15 degree on the upper side and 55 degree on the lower side. The microphone is omnidirectional with noise reduction.

## 5.3 Result
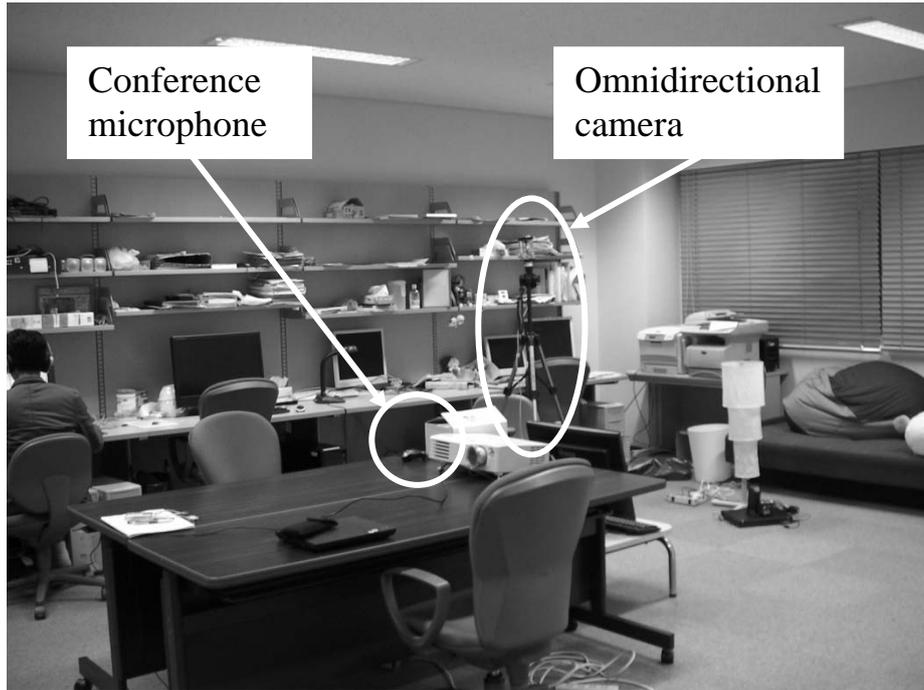
Figure 3: Environment of recording daily informal communication.
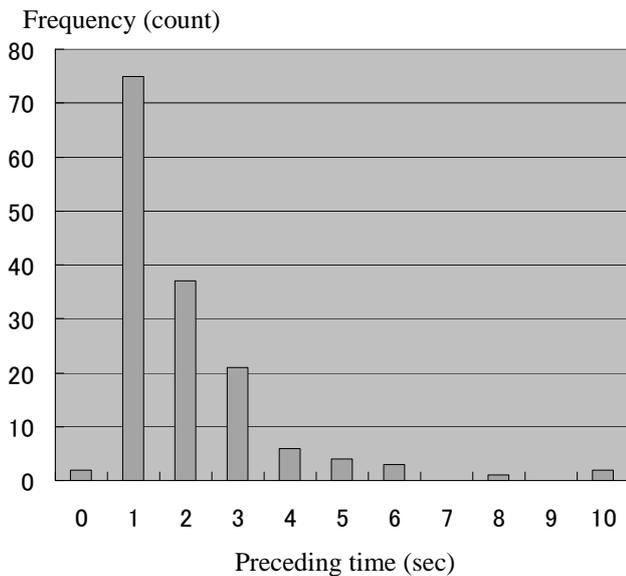


Figure 4: Preceding time of communication behavior before first utterance



Figure 5: Number of participants in a communication

For every communication, time of communication behavior before utterance was clocked. The result is shown in Figure 4. The total number of communication was 151. The average of communication behavior before utterance was 2 seconds. The first utterance came less than 3 second from the beginning in nearly 90% of communication. The longest interval between the beginning of communication and the first utterance was 10 second. It seems to be reasonable that the buffering time of the WCR is 10 second.

Additionally, the number of participants in a communication is shown in Figure 5.

## 6 INITIAL USE OF WEARABLE COMMUNICATION RECORDER

### 6.1 Aim and Method

Table 1: Time between initiation of communication and first utterance of the subject

| communication ID | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| time (sec.) | 1.5 | 0.0 | 4.6 | 3.7 | 3.0 | 4.1 | 2.8 | 0.7 | 2.5 |
| communication ID | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | |
| time (sec.) | 4.0 | 0.0 | 3.2 | 1.2 | 4.4 | 2.0 | 1.8 | 1.4 | |



Frequency (count)
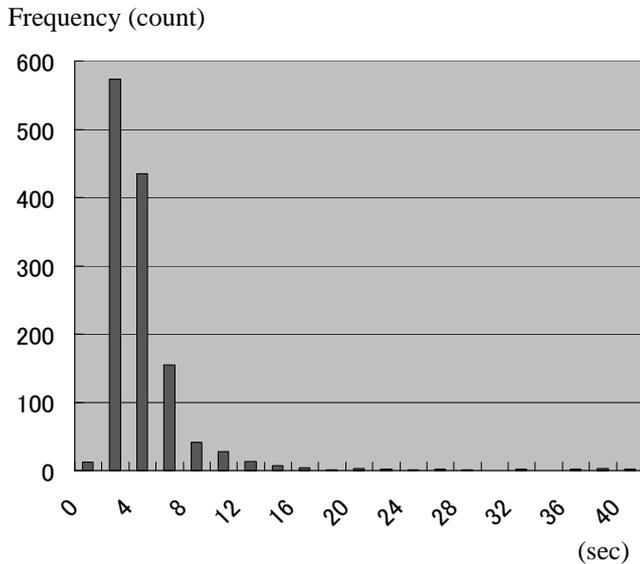
Figure 6: Interval between utterances

To examine feasibility of the system, we used it on trial. Buffering was set to 30 seconds. This means video and audio were recorded from 30 seconds back before the trigger of utterance. From the trial data, it was examined how long initiation of communication come before first utterance of a user. Duration between utterances was also examined to get the clue of completion of communication.

Trial user was one male university student. He wore WCR a day from 10am to 10pm. He communicated with others as same as everyday.

## 6.2 Result

Total recording time was 73 minutes. Speaker and the time of utterance, initiation time of communication by observation were annotated to the video and audio data. This was done by the video annotation software named CIAO.

Communication happened 17 times in the data. Initiation of communication was determined as the point that the subject became aware of the communication target. The time between initiation of communication and the first utterance of the subject is shown in Table 1. Because the number of sample is small, its distribution should not be concluded. But it is suggested that most utterance began within 5 seconds from initiation from the Table 1. It should be noted that all communication were between 2 people. The time is supposed to become longer in the case of communication by more than 3 people. It is suggested that preparation of 30 seconds of buffering time is long enough and probably could be shorter, which is easy to implement.

Then intervals between utterances were examined. Figure 6 is the graph of interval with the scale of 2 seconds for X axis. The most frequent interval was between 2 to 4 seconds. Over 90% was within 8 seconds and over 98% was within 16 seconds. Some data were over 30 seconds but they were about different topics or talking himself. The longest interval in a sequence of conversation was 20.2 seconds. It is suggested that recording can stop about 20 seconds after completion of the last utterance.

Although this initial use is not enough to give detailed data about communication, feasibility of the system was confirmed.

## 7 PERFORMANCE EVALUATION OF WEARABLE COMMUNICATION RECORDER

### 7.1 Aim and Method

Reliability of the automatic recording of WCR was examined in terms of recognition of communication.

A female university student wore the WCR in a room 13 hours in 2days. The room and the setup were the same as in the section 5. All communication in the room was recorded by the omnidirectional camera and the conference microphone, which became the correct answer of communication in the room. The recording taken by the WCR was matched with the answer, thus evaluated the reliability of the WCR. The buffering time was set to 10 seconds.

### 7.2 Result

The number of actual communication was 32. The number of recording of WCR was 39, which means the WCR recognized these as communication. Among these 39, the actual communication was 30.

The WCR could pick up 30 out of 32, i.e. 94 % of all communication, which is known as recall rate. The WCR also sometimes mistakenly recognized the noise event as communication. The rate of actual communication among all recordings was 30 out of 39, i.e. 77 %, which is known as precision rate.

## 8 CONCLUSION

We proposed and implemented a wearable communication recorder (WCR) to get video data of opportunistic communication. Different from other life log recorder or conventional video recording equipment, WCR focuses on recording opportunistic communication efficiently. Video

recording has played an important role in communication analysis but has been applied to limited communication. Significance of WCR is to expand its target. Although current prototype should be improved in some points such as appearance and functions as a wearable computer, it is expected that analysis of opportunistic communication progresses with this type of system.

## REFERENCES

[1] R. S. Fish, R. E. Kraut, R. W Root, and R. E. Rice, Video as a technology for informal communication, Communications of the ACM, Vol.36, No.1, pp.48-61 (1993).

[2] R. W. Root, Design of a multi-media vehicle for social browsing, Proceedings of CSCW'88, pp.25-38 (1988).

[3] K. Ueda, S. Yoshikawa, Y. Den, C. Nagaoka, Y. Ohmoto, and M. Enomoto, Analysis and modeling of conversation, Journal of Japanese Society for Artificial Intelligence, Vol.21, No.2, pp.169-175 (2006). (in Japanese)

[4] C. Nagaoka, and M. Komori, Body movement synchrony in psychotherapeutic counseling: a study using the video-based quantification method, IEICE Transactions, Vol.E91-D, No.6, pp.1634-1640 (2008). (in Japanese)

[5] AIST Press Release, http://www.aist.go.jp/aist_j/press_release/pr2008/pr20081014_2/pr20081014_2.html (2008) (in Japanese)

[6] R. Want, A. Hopper, V. Falcao, and J. Gibbons, The active badge location system, ACM Trans. on Information Systems, Vol.10, No.1, pp.91-102 (1992).

[7] M. Addlesee, R. Curwen, S. Hodges, J. Newman, P. Steggles, A. Ward, and A. Hopper, Implementing a sentient computing system, IEEE Computer, Vol.34, No.8, pp.50-56 (2001).

[8] Y. Sumi, S. Ito, T. Matsuguchi, S. Fels, and K. Mase, Collaborative capturing and interpretation of interactions, Trans. IPSJ, Vol.44, No.11, pp.2628-2637 (2003). (in Japanese)

[9] Tang, J. C., Approaching and Leave-Taking: Negotiating Contact in Computer-Mediated Communication, ACM Trans. on CHI, Vol.14, No.1, Article5 (2007).

# Psychological Type Oriented Adaptive 3D Map in the Mobility

Taizo Miyachi[*], Takashi Aoki[*] , Niikura Tsuneaki[*]

[*] School of Information Science and Technology, Tokai University, Japan
{miyachi, 6add3112}@keyaki.cc.u-tokai.ac.jp

*Abstract* Navigation systems and maps are very useful for a visitor to attain to the destination. However the visitor should find out the target building by himself/herself after the navigation is done. It is sometimes difficult for the visitor to do so in hard environments such that a road has no name and the margin of error of GPS is bigger than the width of the block in a city. Both functions of a map and communication between the user and the map become more important in such hard environments. Adaptive guides corresponding to the psychological preferences make the communication better. In this paper we propose three types of 3D maps based on psychological preference in the mobility and a methodology for adapting to the psychological types of user. 3D maps that supplies spatial information corresponding to the psychological types are also discussed. We evaluated the 3D maps and discuss how the 3D maps works for the users.

*Keywords*: adaptive 3D map, psychological preference type, mobility, navigation system, mashup.

## 1 INTRODUCTION

Navigation systems (NAVI) in car navigation systems, PC, PND and cell phones [7, 8] are useful not only for driving a car but also for walking. However NAVI guides a user only to a place near the destination in Japan since most roads have no name and the margin of error of GPS (Global Positioning System) is sometimes bigger than the width of the block in a city. The user has to discover a target building that has an address in the block. Then communication between human and a map of NAVI becomes very important near the destination. Human has psychological preferences and personality. Adaptive guides for the psychological preferences make the communication easy.

In this paper we propose three psychological types in mobility (a)intuition, (b)rationality and (c)thinking based on MBTI [1] in order to reduce the communication gap. We also propose symbolized 3D map with the altitude corresponding to the psychological types and a method for mashing up them in a Google map [7] in order not to loose the direction on a slanted ground and to find the target building in spite of hard conditions like night etc.

## 2 THE CHANGE OF ENVIRONMENTS AND PSYCHOLOGICAL TYPES IN THE MOBILITY

The environments for ubiquitous computing have been built in order to solve issues based on sensor networks and web sites. Lots kinds of basic information and useful information in a specific condition are utilized for the solution. However the environment changes every moment.

## 2.1 The change of environments and misrecognition

Lots of people live in slanted lands next to flat lands in Japan since Japan has small flat ground (30m times 30m /person) and large mountain lands. There are complex junctions and near-by corners in a small area in the slanted lands and at the foot of a mountain.

Navigation systems (NAVI) have been popular for not only driving a car but also walking etc. However NAVI guides a user only to a place near the destination both in a residential area and in an office area with crowded tall buildings in Japan because of roads with no name and the margin of error of GPS. The user has to discover a target building that has an address in the block by himself/herself.

On the other hand there are many persons who easily loose their ways. People sometimes loose the direction in hard environments. Human's ability for the mobility is lowered in hard environments.

NAVI finds the shortest root for a given destination by tracing connections of the roads even if the route would include the complex junctions with narrow roads. The mobility often becomes hard both at night and in a rainy day.

The major hard environments are (i) to (viii).
(i) Continual small changes: The slanted land with curved roads,
(ii) Daily difficulties in nature: Night
(iii) Natural phenomenon and sudden events: Rain, snow and road works
(iv) Duplication problems. Duplication of address: Many buildings locate in the same address
(v) Similar names: Sato building, Saito building
(vi) View problems. Out of view: An entrance behind the block or behind buildings.
(vii) Psychological effects: Fear, anxiety of getting lost
(viii) Differences with some predictions based on the description of map.

Suitable complementary information should be added so as to adapt for the hard changes.

## 2.2 Psychological types in mobility

Carl Gustav Jung found that human has psychological preferences that influence ways for processing information, making decisions, and perceiving the environments [2].

The communication between a user and a map of NAVI is indispensable in the hard environments after the end of navigation. Human has each character and wavers in unfamiliar situations. Human acts based on his/her true nature in the hard situation. He/she chooses the reliable way for himself/herself to start communication with a map corresponding to his/her psychological preference and disquiet (refer to Fig. 1). He/she also tries to reduce misrecognition both of geometry information in the map and of landmarks in the real route.
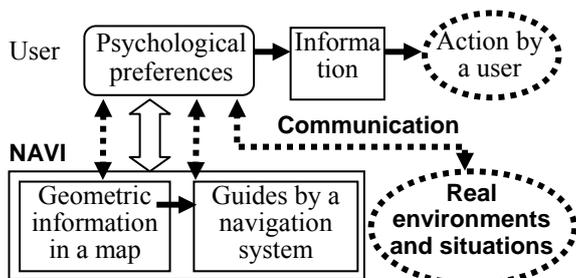


Fig. 1 Communication between a user with psychological preferences and a navigation system

Human can easily understand a message from the others and reduce the communication gap with them if human could consider the differences between psychological preferences each other.

C. G. Jung proposed the existence of two dichotomous pairs of cognitive functions [2]:
- The "rational" (judging) functions: thinking and feeling
- The "irrational" (perceiving) functions: sensing and intuition

The Myers-Briggs Type Indicator (MBTI) assessment is a psychometric questionnaire designed to measure psychological preferences in how people perceive the world and make decisions [1]. These preferences were extrapolated from the typological theories originated by C. G. Jung. The MBTI sorts some of psychological differences into four opposite pairs, or "dichotomies," with a resulting 16 possible psychological types.

We propose three psychological types (a) to (c) for the communication in the mobility based on MBTI in order to allow users to easily understanding both geographic information and the environments and to reduce communication gaps between a user and a navigation system with maps.
(a) Intuition: only key information concerning both the destination and the direction are enough for the communication
(b) Rationality: essentials of geographic information for the mobility are prepared for the communication.
(c) Thinking: detailed information for the mobility in many kinds of case is required for avoiding many kinds of risks.

## 3  ADAPTIVE NAVIGATION BY SYMBOLIZED 3D MAP

People can easily reach near the destination by NAVI. They must find out the target building by themselves with an address in the block in Japan. It becomes harder for them to do so in hard environment like darkness at night.

### 3.1  Adaptive navigation corresponding to psychological preferences

A visitor tries to discover a clue for the right recognition of the surroundings and to get the geographic idea of the target building in the whole area in the hard environments. He/she should also take care of avoiding both dangers and risks that are caused by his/her actions. The visitor chooses the reliable way to acquire useful information in hard environments according to his/her psychological preferences. The psychological preferences are classified into some different types. A psychological preference type of a user often changes into the other type according to user's low capability of consulting maps and the hard environments.

Adaptive guides for psychological preferences make the communication better. NAVI should adjust its type of psychological preference to the type of user (refer to Fig. 2).
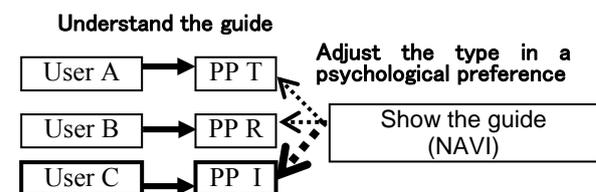


Fig. 2 Psychological preferences for processing based on the ability

NAVI should also adapt for the change of psychological preference type of the user that has changed in a hard environment (refer to Fig. 3). The anxiety of getting lost also changes the type.

User C that can not consult with a map needs the detailed guide and expects to easily recognize the surroundings with landmarks. NAVI should show the user key landmarks and road conditions in order to easily find the landmarks and to map them into the real surroundings.
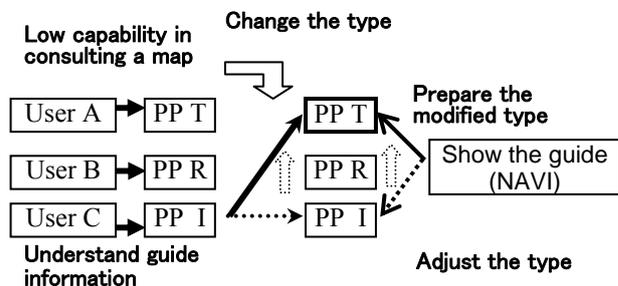


Fig. 3 Preparing navigation for the modified type of Psychological preferences

### 3.2  Three types of simple 3D MAP corresponding to the type of psychological preference

People can understand the geographic information in a map easier than that in a text. Maps are useful much more than navigation in order to find the target building since navigation systems only shows a point near the target building in Japan. People usually use minute maps and aerial photographs (refer to Fig. 4, 5) in order to know a route to the destination. However an intuitive image of 3D structure of installations like an entrance of highway can not be recognized with such minute maps.



**Fig. 4 An example of Google map**

**Fig. 5 An example of aerial photograph in Google map**

3D map in a car navigation system shows intuitive image of 3D structure of an entrance of superhighway. The 3D maps are usually prepared for important spots like intersections and superhighways. Bing map shows buildings in 3D map in Tokyo and major cities (refer to Fig. 6)[9]. However ordinary 3D maps, Bing maps and even Street-view of google maps can not represent the detailed altitude of ground. Curved slopes, steep slopes, stairs by stones and complex roads in the slanted ground often make a visitor loose the direction and enter a dangerous spot. Adaptive guides with 3D maps corresponding to psychological preferences of the user make the communication easier even in such hard environments.

We propose adaptive 3D maps that supply a user with both geographical features and spatial information by symbolized roads with the altitude and warnings corresponding to his/her ability. Lots of people live in the slanted lands in Japan. It is not easy for a visitor to find the target building in such environments. We discuss 3D maps for the slanting land in order to easily reach the target building.

Important geographic information in a map in hard environments is classified into ten categories (1) to (10). (1)direction, (2)distance, (3)landmark, (4)intersection, (5)corner, (6)direction of the slant of land, (7) the degree of inclination of slope, (8)width of road, (9)danger, (10)risk

We propose three types of symbolized 3D maps corresponding to the type of psychological preferences.
(a)3D MAP-intuition: Only key information such as the direction to the destination, a list of corners to change direction, landmarks, and direction of the slant of land are shown (refer to Fig. 7). A higher road is described as a wider road. The lower parts of a down road are described in narrow. The Prefecture (State) roads are surrounded in yellow and major roads are surrounded in     pink.

(b) 3D MAP-rationality: Essentials of geographic information from (1) to (6) for the mobility are prepared for the communication (refer to Fig. 8). The user can get the idea of route with the accurate distance with a few tens meter ruler.
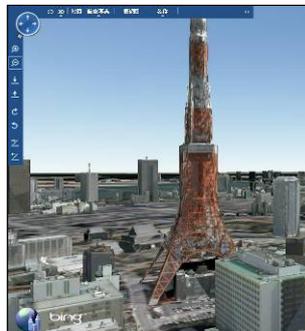


**Fig. 6 An example of 3D map of Bing maps**
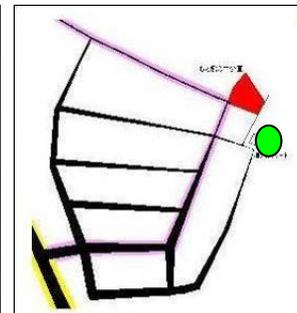
**Fig. 7 A 3D map-intuition corresponding to intuition psycho-logical type**
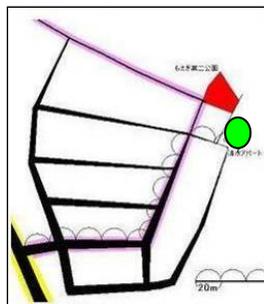


**Fig. 8 3D Map-rationality**
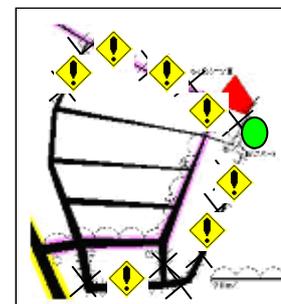
**Fig.9 3D Map-thinking**

(c) 3D MAP-thinking: Essentials of geographic information and warnings in many kinds of cases are shown for the user in order to avoid typical risks and to recover the wrong actions. The information from (1) to (10) is shown in the 3D map as a help information (refer to Fig.9). A user that looses his/her way could avoid the risks by risk icons in the 3D map.

Major functions of the 3D maps for each psychological type in the mobility are (1) to (7).

■3D Map-intuition
(1)Geographical features. The widest road shows the highest road. The wider road is the higher road. A visitor can easily get the intuitive idea of the geographical features with major roads near the destination.

(2)Ensure the direction. A visitor often looses the direction after going down on the curved road in order to discover the target building. He/she can easily get the intuitive idea concerning direction of the slant of land in the 3D maps and can ensure the direction by the feeling of gravitation in the real world. He/she can also make a correspondence between 3D map and the real slope.  A user that can consult a map easily finds a favorite root.

■3D Map-rationality
(3)Near-by corners. NAVI can not show an accurate guide in case that the margin of error of GPS (Global Positioning

System) is bigger than the distance between corners. The visitor can ensure the target corners by twenty meter ruler based on the distance between the related corners even if there would be many corners in a small area (refer to Fig. 8).

(4)Avoid passing the point with accurate distance. It is not easy for a visitor to find the target building if there would be no sign near the target building. A visitor can check the distance from the landmark by the few tens meter ruler and can avoid passing the point in the route (refer to Fig. 8).

■3D Map-thinking
(5)Discovery and recovery of passing the point. A visitor can notice that he/she passed the target point when he/she finds a landmark like a park if he/she would know that the park means passing the target point (refer to Fig.9). Then he/she can start the recovery of passing the target point.

(6)Avoid risks. A visitor must go up the steep slope if he/she would happen to go down on a curved road into a place under the cliff that has one or two entrances of path up to target area. It is not easy for him/her to find the entrance since he/she must choose the right path at a junction on the way back. The visitor could not often notice the existence of the other branch roads at the junction when he/she went down on the curved road. He/she surprises at the branch roads at the junction in the way back. Then it is difficult for him/her to choose the right road. 3D Map-thinking shows users such roads with the risk marks like "×" or "⚠" in order to avoid wandering into the risk area (refer to Fig.9). Visitors can ensure the safe route avoiding the risks.
(7)No sense of direction. Even a visitor who has no sense of direction can finds a route along the edge like a cliff and reaches the target building without getting lost. He/she only go along the edge after he/she reaches a point in the edge (refer to Fig.9 and Fig.14).

# 4 MASH-UPING MINUTE MAPS AND SYMBOLIZED 3D MAPS

Maps are very useful for showing the location of points of interest (POIs) [3]. People usually use a minute map and an aerial photograph for this purpose. Google map has excellent functions for this purpose.

Car navigations and pedestrian navigations are kinds of new applications of maps which need new functions. The views in photo at an intersection allows user to easily find the right route without fails. 3D maps are very useful for represent 3D spatial information in some situations such as showing both the entrance and the right lane at a junction into a superhighway. New types of information for new application should be supplied on demand. The navigation in hard environments is belonging to this type. A 3D map for a slope is a good example.

We propose a mush-up system [5, 6] so as to show 3D maps in an balloon for the destination in a minute map by Google (refer to Fig.10). The user can easily map the roads in the symbolized 3D map to those in the ordinary map since the symbolized map is displayed over the target area.



Fig.10 Mash-up with a minute map and 3D maps (3D Map-intuition)

A user can choose one of three psychological preference types of 3D maps corresponding to his/her psychological preference type that is decided based on both the trip ability of the user and the hard environments.

# 5 EVALUATION

We evaluated the three types of symbolized 3D Maps comparing with Google map and NAVITIME map.

■**Subject**: Subjects are eight men (7 in their 20s one in his 50s) and one woman (in her 20s).

**Test 1. Altitude. Can a user recognize the altitude of slope in either Googlemap or abstracted 3D maps ?**
(Refer to Fig.11 and Fig.12)
Results 1. All subjects could not feel the altitude in a Google map (refer to Fig.11). All subjects except one subject could recognize the altitude in 3D map-intuition (refer to Fig.12). The one subject could not recognize the altitude exactly since there is no exact value of the degrees of an inclination in the map. He requested



Fig.11 An example of Google map

Fig.12 3D map-intuition in the annotation of the destination

additional explanation of the inclination in the 3D map. We found that the map had to display the meanings in it for such a user.

**Test 2. Feature of ground. Can a user get an image of geographic feature of the ground?**
Results 2. All subjects could not get the image of feature of ground in the Google map but could get the image of the feature in the 3D map-intuition (refer to Fig.11, 12).

**Test 3. Distance. Can a user know the accurate distance between corners and intersections?**
Results 3. All subjects could not know the distance in the Google map but could know the accurate distance between corners by the 20 meter ruler in the 3D map- rationality (refer to Fig.13 (c)).

**Test 4. Start a trip. Can a user of the maps start a trip and reach the target building without loosing his/her way?**
Results 4. Six subjects answered that they could start the trip with Google map (refer to Fig.13 (a)). Two of them were confident of reaching the target building since they were good riders either of bicycle or car and had experiences of many trips. The other subjects did not have any confidence of reaching the building. Three subjects did not expect start with the Google map. They needed more geographic information for the start.

All subjects answered "yes" for starting the trip with 3D map-rationality. Three subjects were full of confidence that they could reach the building since they could know the accurate distance between corners by the 20 meter ruler (refer to Fig.13 (c)).

**Test 5. Notice risks and have fear of dangers**
Results 5. Three subjects could feel relieved about ensuring areas with risks in the 3D map-thinking (refer to Fig.13 (d)). One subject did not need information concerning danger since he was a good rider of bicycle and could go anywhere without loosing his ways.
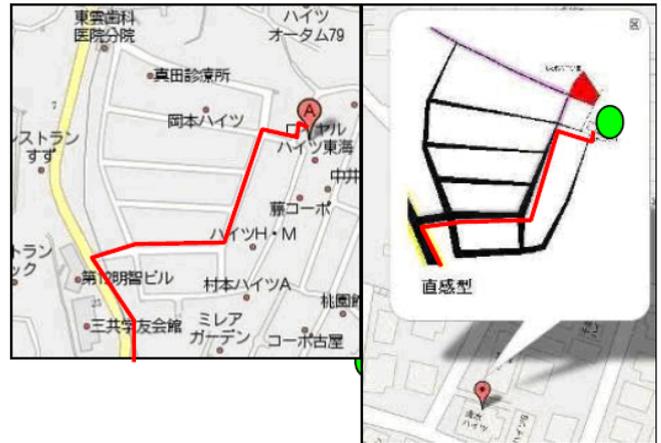


(a) Route guide in Google map

(b) Mash-up with a minute map and **3D Map-intuition**
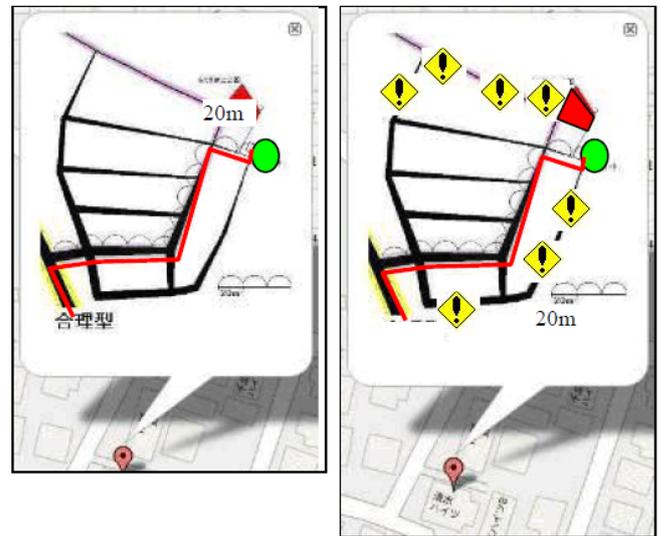


(c) Mash-up with a minute map and 3D Map-rationality

(d) Mash-up with a minute map and **3D Map-thinking**

Fig.13: 3D maps corresponding to psychological types

**Test 6. Avoid getting lost by going along the edge**
Results 6. One wise subject that always looses his way proposed a route that utilizes the edge like a cliff in spite that there are marks of risks on the route (refer to Fig.14). He expected to turn right at the first corner and to go to the edge since he only kept going along the edge to the destination after he reached the edge.

■ Discussion. We propose three types of 3D map representation to provide required and sufficient information for a visitor based on the psychological preference such as intuition, rationality and thinking. Information to be displayed is controlled to tell the users only the truly required information since a visitor does not need too much detailed information. A visitor who easily looses his/her way usually is a user of 3D Map-thinking. We found that the wise subject in Test 6 is a kind of special user of 3D Map-

thinking since he basically believes that he usually looses his way. 3D Map-thinking should provide both the shortest route and
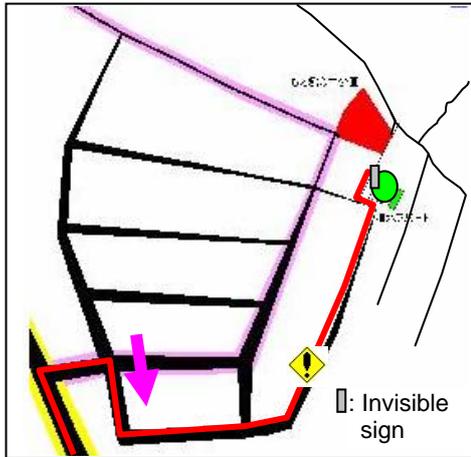


Fig.14 A safe route along the edge without loosing ways

of the controlled display is the sign of the target the safe route for such a special visitor. The other reason apartment could not be seen from the safe route. 3D Map-thinking should not display the sign with the safe route. The special visitor can not find the sign and passes over the apartment. Unfortunately he un-intentionally goes down into a curved narrow road and notices that he lost his way after a while. He can not recognize the right road to return on his way back at the intersection that he passed.

Most people that choose the shorter route avoid the safe route in the wrong direction. They feel dread when he entered into a narrow side road that is also a curved road if they would choose this road. Many of them need not such information in 3D map-intuition and 3D map-rationality.

## 6. CONCLUSION

A visitor should find out the target building even in the hard environments at the destination where a navigation system brings the visitor. He/she communicates with maps based on his/her psychological preferences then. We propose three types of psychological preferences in the mobility (a) intuition, (b) rationality, and (c) thinking based on MBTI in order to allow the user to easily acquire key information for the mobility. A visitor feels fear and easily changes into the different type when he looses the direction either in the night or at the curved slope etc. We also propose three types of adaptive symbolized 3D maps corresponding to the three types of psychological preferences in order to supplies the user with the altitude, the feature of ground like a slope, near-by corners and dangers etc.

We tested the adaptive 3D maps for eight subjects. All subjects could get an image of the slanted slope and the altitude in the 3D maps. All subjects also answered that they could start the trip with 3D map-rationality. Three subjects were full of confidence that they could reach the building since they could know the accurate distance between corners by the 20 meter ruler. Three subjects could feel relieved about ensuring areas with risks in the 3D map-thinking. We

could ensure the usefulness of the 3D maps corresponding to the three types of psychological preferences in the mobility.

We would like to investigate on navigation systems that can flexibly adapt for a variety of change of psychological preferences in future.

## REFERENCES

[1] I. B. Myers, and P. B. Myers, Gifts Differing: Understanding Personality Type. Mountain View, CA: Davies-Black Publishing. (1980, 1995)

[2] Jung, Carl Gustav. Psychological Types, Collected Works of C.G. Jung, Volume 6. Princeton University Press, ISBN 0-691-09774. (1971)

[3] Luca Chittaro, Visualizing Information on Mobile Devices, Computer (published by the IEEE Computer Society), 39(3): pp. 40-45, 2006.

[4] B.B. Bederson et al., "DateLens: A Fisheye Calendar Interface for PDAs," ACM Trans. Computer-Human Interaction, vol. 11, no.1, 2004, pp. 90-119.

[5] Naoyuki Nomura, Comparing Web API, "http://www.api-match.com/search/index?/anguage =en," 2008.

[6] Michael Bolin, End-User Programming for the Web, MIT MS thesis, 2005

[7] Google maps, http://www.google.com/maps

[8] NAVITIME, http://www.navitime.com/

[9] Bing maps, http://www.bing.com/maps/

# A Vehicle Information Transmission Method by Inter-Vehicle Communication and its Application to Traffic Jam Detection

Kenta Ohmi*, Takeshi Nishimura**, Jun Sawamoto***, Hisao Koizumi****

*Information-Technology Promotion Agency (IPA), Japan
**Mitsubishi Electric Corporation, Japan
***Faculty of Software and information Science, Iwate Prefectural University, Japan
****Department of Computers and Systems Engineering, Tokyo Denki University, Japan
sawamoto@iwate-pu.ac.jp, koizumi@k.dendai.ac.jp

*Abstract* - Currently, car navigation systems are advancing in the field of Intelligent Transport Systems (ITS), thereby advancing traffic information systems. Further reduction of economic losses arising from traffic jams is anticipated through inclusion of traffic information systems. Vehicle Information and Communication Systems (VICS), VICS probe, and Internavi Floating Car Systems with telematics have therefore been researched and developed. This study specifically examined multi-hop inter-vehicular communication and the vehicle information acquired by it. We designed a traffic information system that improves accuracy and real-time features of the acquired traffic information without a communication-based VICS center, traffic information center, or server: each vehicle acquires other vehicles' information using real-time inter-vehicular communication. In this paper, we propose a method of vehicle information transmission using inter-vehicular communication. We also attempt to detect traffic jam situation using vehicle information that each vehicle receives.

*Keywords*: ITS, Car navigation system, traffic information, VICS, telematics.

## 1 INTRODUCTION

Currently, there are nine development fields of intelligent transport systems, among which is the field of highly advanced car navigation systems [1]. Of such systems, Vehicle Information and Communication Systems (VICS) [2] have difficulty acquiring traffic information correctly except for information related to links (roads) where sensors are emplaced. Therefore, a method that is currently under consideration is the use of a VICS probe [3] to improve the level of traffic information accuracy and traffic volume measurement by collecting probe data that the vehicle can sense using such bi-directional communication equipment as a light beacon or an electric wave beacon on roads with no sensors in place. Any such traffic information system gathers information once at an information center. It then selects and transmits necessary information to other vehicles, in which sense we can call it an infrastructure-dependent system. In contrast, infrastructure-independent systems exist, for which prior studies to utilize inter-vehicular communication methods have been made. One such system is the Advanced Safety Vehicle (ASV), which transmits vehicle location information from one vehicle to another using inter-vehicular communication. Then all available short-distant vehicle locations are displayed on the car navigation system screen to notify the driver [4]. Thereby, inter-vehicular communication is used to transmit the real-time information. It is considered useful for multi-hop communication utilizing Mobile Ad-hoc NETwork (MANET) technology [5].

Traffic View [6] is a technique for a vehicle to obtain information of all other vehicles moving on a straight road such as a highway. It transmits the collected vehicle information to trailing vehicles using one-hop communication at a constant transmission timing that is preset for all related vehicles. Another technique, RMDP [7], reduces transmission message collisions and improves transmission rates by determining every interval as to when to transmit next based on the amount of data received during past transmission intervals of the vehicle and the number of error messages caused by data collisions. From such a study of transmitting vehicle information using inter-vehicular transmission techniques, tradeoffs are understood as the most important factor to consider, the resolution of which yields the right message size of the vehicle information for every usage, the transmission interval, the transmission distance, and the real-time necessity of transmission.

One matter to note is that when VICS tries to measure the traffic flow on every link to determine possible traffic jams with information from sensors set along the road and vehicle driving times on the road, it normally, and invariably, takes some time to do so, thereby producing some discrepancies in measurement from the actual traffic situation. To avoid that situation, we proposed a traffic information system to notify drivers of traffic information on a real-time basis. The system analyzes the surrounding traffic status by allowing every vehicle to send traffic information to surrounding vehicles at a constant interval within seconds [8, 9].

In the study described above, acquisition of information related to all surrounding vehicles is done using a flooding technique [10], which is recognized as a broadcast type transmission technique. However, one problem is that a broadcast storm phenomenon [11] can be produced by transmission flooding, which results in failure of traffic information transmission. In such a storm, the transmission interval used by every vehicle becomes shorter and the amount of communication equipment that is used increases in scale. A study of Location Aided Routing (LAR) [12] has been made to efficiently construct a routing table based on location information of every mobile node, but constructing a routing table engenders difficulty because the vehicles

concerned move so rapidly that communication frequently gets cut off.

In this paper, we propose a vehicle information transmission method in line with improving its transmission rate along with flooding to help improve the accuracy level of traffic flow measurement. In addition, we seek a means to evaluate a possible traffic jam situation based on information received from each vehicle, and show how it could be used in the real world. The former comprises two transmission methods, one of which is a means to delete messages to transmit the previously received vehicle information in order to reduce the transmission load. Another is a means to reduce the broadcast message transmission load by forming a group of vehicles for link-by-link messaging, sending only a representative message to every one of such link groups. The latter method judges whether a traffic jam is possibly occurring or not on every link by computing the weight level of every vehicle on the link depending on where it is located.

Section 2 proposes a vehicle information transmission method while improving the flooding technique. Section 3 presents an evaluation of such a proposed vehicle information transmission method using a simulation technique. Finally section 4 presents examination of how such a proposed method can detect a possible traffic jam situation from available vehicle information as that used in the real world.

# 2 VEHICLE INFORMATION TRANSMISSION BASED ON INTER-VEHICULAR COMMUNICATION

## 2.1 Vehicle Information Transmission

Vehicle information sent and transmitted through inter-vehicular communication includes a vehicle identification code, location, speed, acceleration, moving direction, time of sending the information, and a link identification code. Every vehicle is supposed to be equipped with a GPS receiver, a car navigation system, and a wireless inter-vehicle communication device (IEEE802.11x or similar device). Using a GPS alone, the measurement of the vehicle location can be known within several meters, but the current car navigation system uses a map-matching technique, the error can be reduced further to sub-meter-order as long as the vehicle is moving on a road. A GPS can be used not only for measuring the vehicle location; it can also synchronize the physical clock of the vehicle's computer, based on time information sent from the GPS satellite, with precision of a few microseconds. Current technology levels of the vehicle can already sense the speed, acceleration, and moving direction of a vehicle. Furthermore, when a message is broadcast, it includes information of the vehicle from which the transmission is made. That information can be useful to delete the message by the receiving vehicle, as described later. Based on the definition of these pieces of vehicle information, Table 1 presents some actual examples of vehicle information held in memory.

Each piece of vehicle information would be transmitted with a vehicle identifier (ID). Every time a message is received from any surrounding vehicle, vehicle information with the same identifier would be constantly updated to include up-to-date vehicle information.

Table 1: Examples of vehicle information

| ID | speed | acceleration | direction | latitude | longitude | time | Link ID |
|---|---|---|---|---|---|---|---|
| 10 | 30 | 5 | 2.45 | 25.44 | 136.44 | 120030 | 1732 |
| 11 | 45 | 2.5 | 3.43 | 25.43 | 136.45 | 120040 | 4723 |
| 15 | 20 | -5 | 0.03 | 25.45 | 136.39 | 120115 | 9612 |

## 2.2 Method of Deleting Vehicle Information Messages

In our method, it is not assumed that each vehicle acquires information of any specific vehicle per se, but we basically use a broadcasting type of communication such as a flooding method to acquire surrounding vehicles' information. However, when a vehicle regularly transmits its vehicle information messages to all other vehicles using a flooding method, if the transmission intervals are short, then the associated message transmission would fail because some messages will collide. Therefore, in our study, we propose a method of deleting broadcast transmission messages while predicting other vehicle locations based on basic vehicle information such as speed, acceleration, and the moving direction. We call this the message deletion method.

The GPS receiver can not only measure its location; it can also synchronize the physical clock of the vehicle with the time the GPS satellite holds. In other words, the time difference of the messages sent and received among vehicles becomes the time spent for communication, which means that other vehicle locations can be predicted by acquiring information through inter-vehicular communication; location, speed, acceleration, moving direction, and the time that the other vehicles acquire using the installed sensor. Vehicle $V_n$ sends its vehicle information to vehicle $V_l$ before $V_m$ does. Furthermore, location information, $x_n$, $y_n$, of vehicle $V_n$ acquired from the GPS in addition to its speed, $v_n$ [km/h], its acceleration, $a_n$ [km/h/s], the time that message n is sent, $S_n$ [ms], and the time $R_m$ [ms] at which vehicle $V_m$ receives message m. Then the moving distance $l_n$ [m] of vehicle $V_n$ from the perspective of vehicle $V_l$ can be expressed as follows.

$$l_n = \frac{v_n(R_m - S_n) + \frac{1}{2}a_n\frac{(R_m - S_n)^2}{1000}}{3600} \qquad (1)$$

Furthermore, with the y axis for the longitudinal line, the x axis for the latitudinal line, and $\theta_n$ [rad] for the moving direction of the vehicle $V_n$, the predicted location, $X_n$ and $Y_n$, after moving a distance of $l_n$ [m] can be expressed as shown below.

$$X_n = x_n + l_n \cos\theta_n \qquad (2)$$
$$Y_n = y_n + l_n \sin\theta_n$$

An important rule is that vehicle $V_l$ does not broadcast the message it receives from vehicle $V_n$ if the information of all vehicles meets the following conditions: information of vehicles $V_n$ and $V_m$, which are within the range of one hop of each other is received and retained; the predicted location of the vehicle $V_n$, which has sent the broadcast message is $X_n$ and $Y_n$; the predicted location of the vehicle $V_m$ is $X_m$ and $Y_m$; the current location of vehicle $V_l$ that judges the transmission is $x_l$ and $y_l$; and the communication range of the communication equipment installed in each vehicle extends as far as $\alpha$ [m].

$$\sqrt{(X_m - X_n)^2 + (Y_m - Y_n)^2} \le \alpha \; or \sqrt{(x_l - X_m)^2 + (y_l - Y_m)^2} > \alpha \cdots (3)$$

Fig. 1 portrays an actual example using vehicles $V_1$, $V_2$, and $V_3$. It presents a case in which, if vehicle $V_2$ has received messages from vehicles $V_3$ and $V_1$, then $V_2$ analyzes the message sent from $V_1$ and does not make a broadcast transmission when it can predict that $V_3$ will remain within the transmission range for $V_1$.

In addition, because no location can be available for those vehicles from which no vehicle information has been received, it is assumed that there would be no surrounding vehicles: consequently, no broadcast transmission is made.
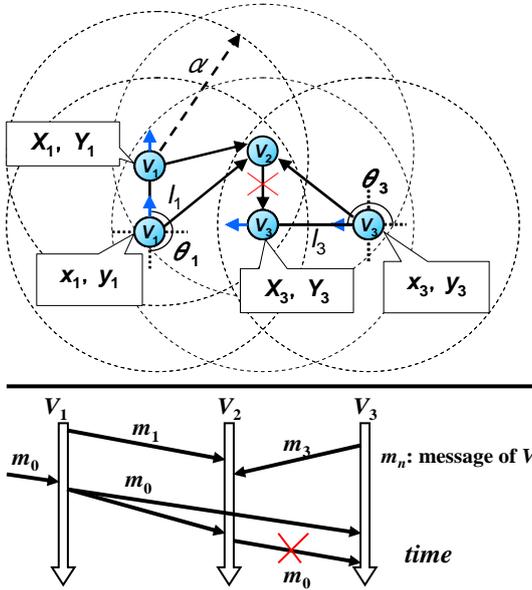


Figure 1: Example of message deletion method.

In the method we propose, the location of other vehicles is predicted based only on vehicle information that has been received already. Based on that information, whether a broadcast transmission is made or not is determined.

Salient differences between LAR and the proposed method are the following.

- The proposed method incorporates acceleration.

- LAR determines and notifies others of the area by examining locations of the sender and receiver, but the proposed method determines whether to transmit or not every time the transmission is made by examining the state of surrounding vehicles.
- LAR considers locations of the sender and receiver only, but the proposed method examines relative locations of all three vehicles.

## 2.3 Information Transmission using a Link Group

To improve the vehicle information transmission efficiency, we propose, in addition to the message deletion method described in section 2.2, a means to choose a representative vehicle for every link and for such a representative vehicle to send the collected vehicle information to surrounding links with its signal range expanded. Herein, we call this proposed method the link group method.

Fig. 2 presents an example of expanding the communication range around the representative vehicle for the link.



▶ representative vehicle    ⌇ representative vehicle's signal range
▷ ordinary vehicle    ⌇ ordinary vehicle's signal range
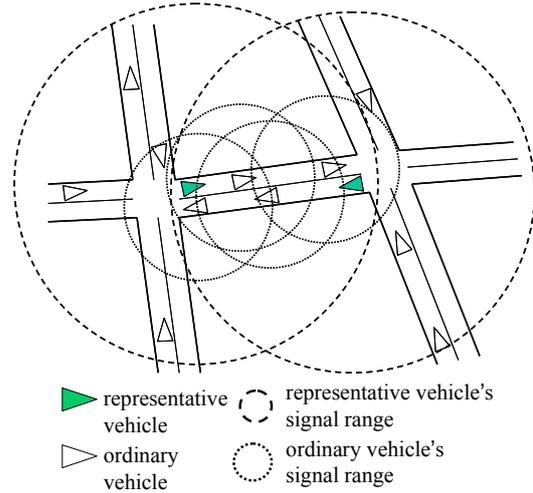
Figure 2: Example of the communication range of a representative vehicle and an ordinary vehicle

On every link, vehicles are classified into two categories––ordinary vehicles and representative vehicles. Then a representative vehicle is defined as one that has newly entered the associated link. A vehicle, when it enters the link, defines itself as a representative vehicle, and sends a message forward to vehicles ahead requesting a change of the representative vehicle role. The representative vehicle gathers vehicle information of all vehicles on the link and creates a link representative message for it. The representative vehicle sets the transmission time using variables obtained based on the identifier of each link and available information about its moving direction, and sends the link representative message regularly. The representative vehicle receives a message from any vehicle behind on the link requesting a representative role change; it then gives the right to that role to that vehicle. Additionally, the

representative message is sent based on the flooding technique, and would be transmitted further from one to the next by other representative vehicles. A link is defined here as a road section from one intersection to another.

Other vehicles, when they receive messages from ordinary vehicles or the representative vehicles of other links, retain that vehicle information, but do not broadcast it. Furthermore, when a new representative vehicle sends a message forward regarding the representative vehicle change, that signal might not reach it if the link is too long. In that case it might be necessary that there be more than one representative vehicle.

# 3 SIMULATION OF VEHICLE INFORMATION TRANSMISSION

In this paper, we use a prototype car navigation system to simulate and evaluate the manner in which the vehicle information would be actually transmitted as we propose.

## 3.1 Simulation method

For simulation of vehicle information transmission, we examine a straight road on a digital map placing the moving vehicles on the road. We evaluate the transmission efficiency while changing the communication range size intentionally. The following describes all simulations.

(1)    Setup of simulation for communication
- The flooding method, the message deletion method, and the link group method will be evaluated.
- The messages sent are actually broadcast ones. Therefore, no ACK process of CSMA/CA will be considered to take place. Detection of message collisions will be made on the message receiving side. We assume that it takes 1 ms to send a vehicle information message. Our model is therefore a simple one assuming that every vehicle uses 1 ms of time during communication when it receives or sends data. The time interval for each vehicle to send data is fixed at 1 s.
- For the flooding method and the message deletion method, the communication ranges from each vehicle will be varied starting from 100 m to 1000 m by 100 m steps.
- For the link group method, the communication ranges that the ordinary vehicles or representative vehicles use will be varied starting from 100 m up to 1000 m in 100 m steps. However, measurements are made only in the setup in which the communication ranges by ordinary vehicles are always smaller than those by representative vehicles.
- Measurements for each method will be made for 10 s. Measurements will be made vehicle-by-vehicle to the average number of transmitted broadcast messages, the average number of received messages, the average vehicle information transmission rate, and the average number of vehicle information updates. The number of vehicle information updates signifies the times the vehicle information received from other vehicles is saved for a new message, or updated for existing messages. The vehicle information transmission rate is

the number of vehicle information updates divided by the total number of messages transmitted by all associated vehicles. No message will be destroyed irrespective of the distance from which they come, but all information of vehicles on the roads will be received and considered.

Fig. 3 presents a simulation model used for messages in sending and receiving. Each vehicle would move during each 1 ms of time, and finishes in sending a message during any one of the 1 ms intervals. The shaded area shows a message collision occurring. In the diagram, messages $m_k$ and $m_m$ collide, for example. Whether the messages collide or not is judged on the message receiving side, but the vehicle, when it sends a message, happening to receive a message from another vehicle at the same time would understand that a message collision had occurred.
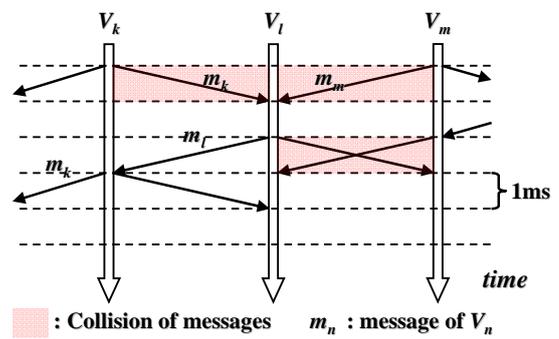


Figure 3: A simulation model used for message sending and receiving

(2)    Simulation setup for vehicle movement
- The straight road we use is assumed as 1 km long having one lane in each direction. On the straight road, we place 200 vehicles randomly, making them run in both directions on the road.
- The actual digital map is assumed to be of Kawagoe city in Saitama prefecture, so the vehicles move randomly on that map. Kawagoe city is as wide as about 20 km × 20 km; that is expressed on the digital map. On it, 3000 vehicles are placed randomly.
- For both Kawagoe city and the straight road, the vehicle movement; the vehicle acceleration is expressed based on the Optimal Velocity (OV) model [13]. The maximum speed of vehicles is 60 km/h for both Kawagoe city and the straight road; the minimum is 0 km/h.

Fig. 4 portrays a screen image of this prototype car navigation system. This car navigation system runs on Linux base, and can use shape file map data [14]. At present, the map data of Saitama prefecture are put in the system. It has such functionalities as present position representation, facility location searching, map zoom in and out, and route searching. Fig. 5 shows a digital map of Kawagoe city and a chart of a vehicle positioning arrangement, both of which are products from the prototype car navigation system. This digital map includes 11,077 links.
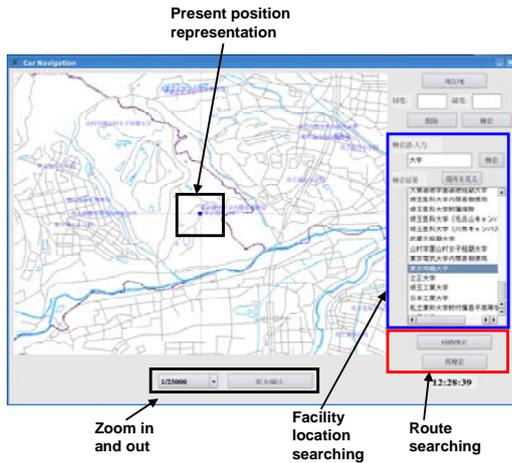
Figure 4: Execution of prototype car navigation
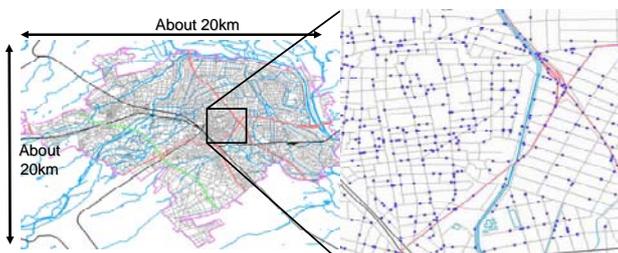


Figure 5: Digital map of Kawagoe city and a chart of a vehicle positioning arrangement

## 3.2 Results from Simulation

Fig. 6 shows the average number of messages sent and received in the straight road environment when the communication range size was varied from 100 m to 1000 m for the flooding method and the message deletion method. The message deletion method shows that the number of messages that are sent is decreasing, although those received increase as the size of the communication range increases.

Fig. 7 shows the average vehicle information transmission rate and the average number of vehicle information updates. The message deletion method shows that both the number of updates and the transmission rate are increasing as the communication range size grows.

Fig. 8 shows the average number of messages sent and received in Kawagoe city when the communication range size was varied from 100 m to 1000 m for the flooding method, the message deletion method, and the link group method. We find that the number of messages sent and received increases more for the link group method than for the flooding method or the message deletion method.

Fig. 9 shows the average vehicle information transmission rate and the average number of vehicle information updates. We find also that the message transmission rate and the update frequency are improved more for the link group method than for the other two methods.
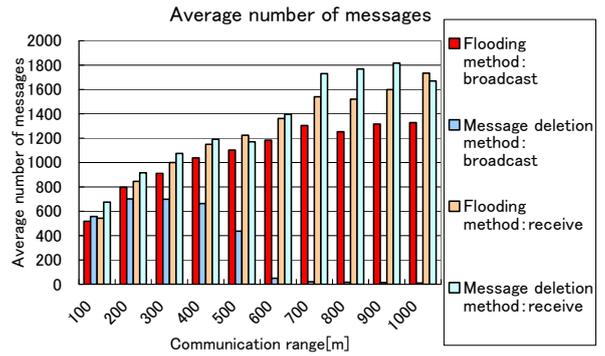


Figure 6: Average number of messages sent and received in a straight road
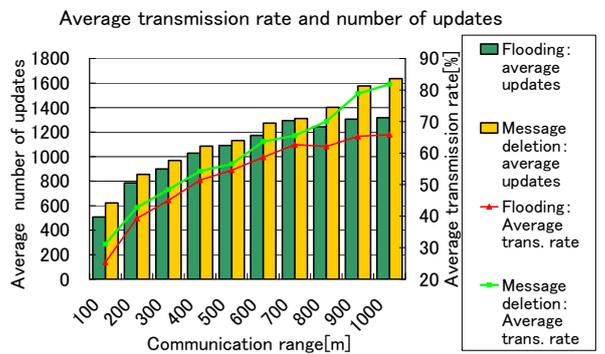


Figure 7: Diffusion ratio of messages and vehicle information update frequency in a straight road
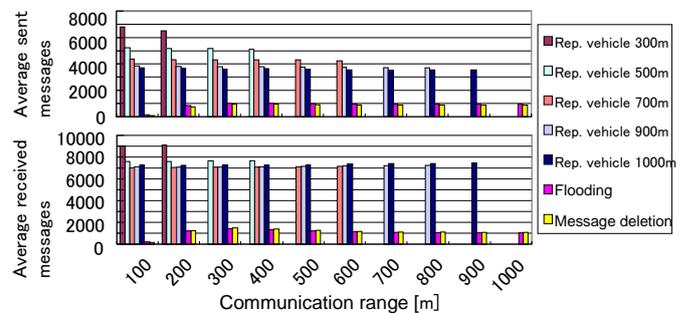


Figure 8: Number of sent and received messages in Kawagoe City
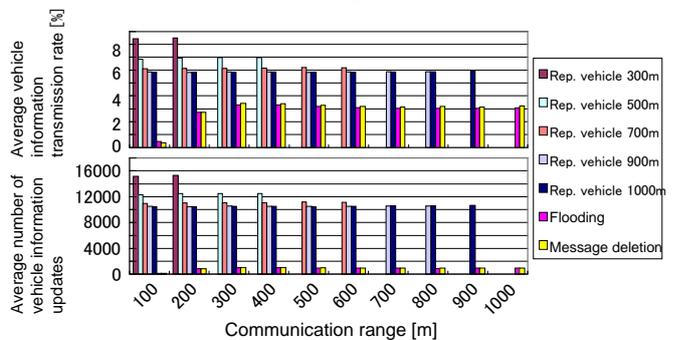


Figure 9: Average vehicle information transmission rate and the average number of vehicle information updates frequency in Kawagoe City

## 3.3 Considerations

We find from Fig. 6 that the number of received messages increases as the communication range increases on the 1 km straight road using the message deletion method, although the number of sent messages decreases. That is true because the range in which the multi-hop transmission had to be made became so limited as the communication range increased that the number of wasted broadcast messages decreased to a great extent in the message deletion method. This concurs with the fact that the average number of message updates was improved so that the message transmission rate improved by 5–15% more compared to the flooding method.

In the simulation for Kawagoe city, the message transmission rate improved when the message deletion method was used alone compared to that obtained when using the flooding method, but that improvement was not as great as that for the simulation using the straight road. However, when it comes to the link group method, its efficiency was certainly recognized. As Fig. 8 shows, the number of sending messages increased by 3000–5000, and the number receiving messages by 6000–7000. That improvement occurred because the number of message collisions decreased as much as the number of messages transmitted from all vehicles had decreased. As a consequence, the number of messages sent and received increased. An additional finding is that the number of sent messages decreases as the communication range covered by a representative vehicle expands, but the number of received messages does not decrease as much as the decrease for the sending ones. That is true because the message deletion method is used in parallel. A noteworthy point is that the greatest number of receiving messages was obtained with the communication range set to 200 [m] for ordinary vehicles, and 300 [m] for representative vehicles. The result was different for the case of using a straight road, in which case we had increasingly better results as the communication range was expanded.

Fig. 9 shows that the number of message updates increased even to 10 times that when flooding was used. Additionally, results show that, at maximum, the message transmission rate is more than twice as much as that for the flooding, which underscores the efficiency of the link group method.

The simulation this time showed the best results of all with the setting of the communication range of 200 [m] for ordinary vehicles and 300 [m] for representative vehicles. However, the message transmission rate for those was as low as 8.5%: the vehicle information transmission method that we propose has difficulty obtaining vehicle information in the range of 20 km × 20 km. Consequently, we would have to narrow down the simulation range to find the most efficient range size for our proposed methods.

## 4 APPLYING VEHICLE INFORMATION TO TRAFFIC JAM DETERMINATION

## 4.1 Weighted Method for Traffic Jam Determination

We next propose a means to detect and evaluate a possible traffic jam situation based on transmitted and available vehicle information. In general, the traffic flow volume increases as the vehicle density on the link increases. However, beyond a certain level of vehicle density on the link, the traffic flow volume begins to decrease. Given the ability to obtain vehicle information from all vehicles, it would be possible to measure the vehicle density on the link based on the position of every vehicle on it. Thereby, we would be able to judge whether a traffic jam is occurring or not. However, not all vehicles have inter-vehicular communication equipment installed. Moreover, obtaining all necessary information through inter-vehicular communication is difficult. We therefore propose a method to detect possible traffic jams based on vehicle information; the vehicle's speed, acceleration, direction, and position on the road obtained from vehicles with inter-vehicular communication equipment while considering other vehicles that have no such equipment.

We first define a vehicle as a traffic-jam causing element or not in the following manner. Based on traffic information, the vehicle speed in traffic jam is supposed to be 10 km/h or less. Moreover, we infer that the vehicle acceleration would decrease in a traffic jam. We therefore define the vehicle as a traffic-jam causing element when its speed is 10 km/h or less and its acceleration 5 km/h/s or less. Additionally, we consider that some vehicles are in a halted state near the link exit area, where there are normally traffic lights, and that some other vehicles are turning right or left. Considering that fact, we weight the vehicle information based on where the vehicle is located on the link. We assign less weight to those vehicles that are about to leave the link in the forward direction; we assign more weight to those near the link entrance area. Such a weighting parameter β is defined as follows.

$$\beta = -P + 1 \cdots (4)$$

$$P = \frac{Location\quad of\ Vehicle\quad [m]}{Length\quad of\ Link\quad [m]} \cdots (5)$$
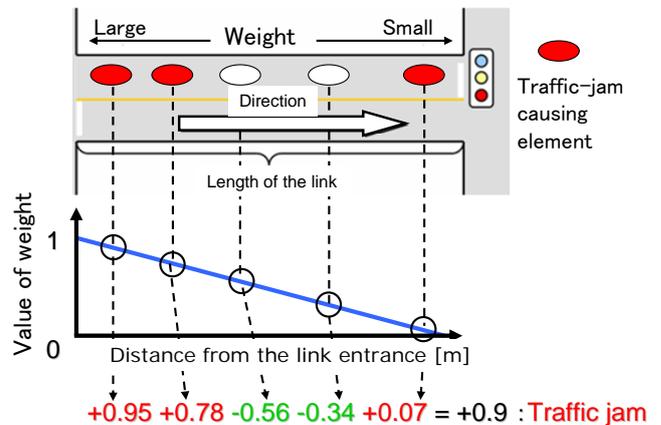


Figure 10: Example of weight according to the vehicle position in the link

We have a positive number for traffic-jam causing elements (with a maximum value of 1), and a negative number for non-traffic jam causing elements (with a minimum value of -1). Summing them up for every link, if the total is negative, we infer that a traffic jam is not occurring. In contrast, if it is positive, we infer that a traffic jam is occurring. With a positive number of summed values, if reached, we can use it to know how heavily a traffic jam is occurring. The closer it is to zero (0), the lower the probability that a traffic jam occurring. The larger it becomes, the greater the chance that we have a traffic jam. In Fig. 10, we have assigned a positive value weight to the traffic-jam causing element, and a negative value to non-traffic jam causing element. Their sum was 0.9. We therefore infer that a traffic jam is occurring on this link.

## 4.2 Simulation for Traffic Jam Determination

### 4.2.1    Simulation method

For the simulation we made to measure the traffic flow volume, we constructed a traffic flow simulator and drew a curve of traffic flow (Q) versus vehicle density (K). Using this Q–K curve, we make a comparison to elucidate the degree to which our proposed method agrees with simulation results in terms of whether a traffic jam is occurring or not.

Generally regarding the traffic flow simulator, it includes a macro model and a micro model. The individual behavior of each vehicle is examined in our proposed system described herein. Therefore, we constructed a traffic flow simulator using an OV (Optimal Velocity Model) [13] model for the micro model. The OV model defines the vehicle speed based on the distance from the leading vehicle. Using it, we made the following settings to do the simulation.

- We had 1 km length of straight road for the traffic flow to run on with 1 lane in one direction using that one lane only. Fig. 11 shows the screen image we used when we executed the simulation.
- We assumed a 17 [m] inter-vehicular distance, below which the speed of the vehicle would have to be decreased to avoid collision, in addition to using 16.7 [m/s] of maximum speed (about 60 km/h) and sensitivity of 1.0.
- We put a vehicle flow onto the road in either a 1, 2, … ,10 s interval in the straight line direction, and determined its entry interval randomly every 10 min.
- We set the traffic flow measurement location at 250 [m] from the link entrance point, and measured the number of vehicles passing during the 5-minute period in addition to the vehicle density.
- We had traffic lights at the road exit location, as normally exist on actual roads, and changed lights from red to green or oppositely every 30 s. The vehicles stopped for traffic lights.
- We continued the simulation for 100 [hr] (with 1200 samples). We made the vehicles move every 10 ms.

Fig. 12 shows the Q–K curve resulting from the simulation. Using the Q–K curve drawn based on the OV

model and knowing whether we had a free flow or a congested flow, while judging whether we had a traffic jam or not using the vehicle information available at every 5-minute interval, we measured the traffic jam agreement ratio between the two. Along with the settings we made as shown above, we produced measurements with the following settings.

- Considering the number of vehicles with inter-vehicular communication equipment installed, we vary the vehicle information acquisition ratio on the link by 10% increasing it by 10% steps gradually to 100%. Then we examine the situation and measure the traffic jam detection agreement ratio when using our proposed method for every vehicle-information acquisition ratio as established above.
- We measure the traffic flow based on our proposed method using vehicle information available at every 5-min interval. Then we compare the traffic jam situation with that available from the Q–K curve.

Moreover, along with the weighted function of the proposed method, we compare it using the step function (with a fixed value of β=1) and function curve shown below.

$$\beta = \frac{(\cos(P \times \pi) + 1)}{2} \cdots (6)$$
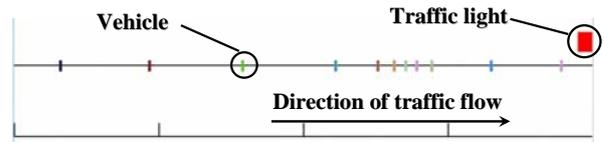


Figure 11: Screen image of simulation execution
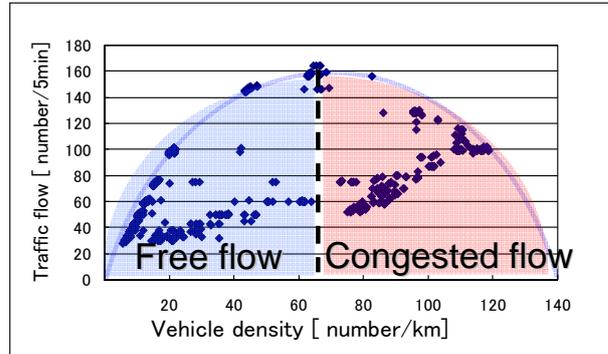


Figure 12: Q–K curve with the OV model

### 4.2.2    Simulation results

Fig. 13 presents evaluation results for the traffic jam detection agreement ratio obtained from the OV model that we used. Using the simulation for traffic jam detection, the results as shown in Fig. 13 reflect that measurement of the traffic flow volume based on its instantaneous state using the vehicle information available provides greater than 80% agreement, even when using a low vehicle information acquisition ratio. Furthermore, our proposed method, in which we put higher weight to the vehicles near the link entry point, yields a better agreement ratio by 5% compared to that using the step function. Moreover, results shows that

more precise measurement results can be obtained when using a simple linear function rather than using a cosine function. These results show the possibility of judging whether a traffic jam was occurring or not if the vehicle information acquisition ratio was greater than 20%.
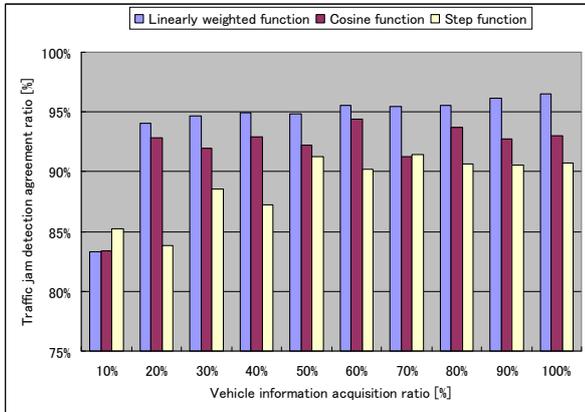


Figure 13: Traffic jam detection agreement ratio

## 5 CONCLUSION

In this paper, we proposed a method of using inter-vehicular communication to transmit vehicle information that every vehicle would measure. Based on the flooding method of transmitting vehicle information, we proposed a means to reduce the transmitted message volume using already received vehicle information. Moreover, we proposed a means to define a representative vehicle for every link and to send the vehicle information collected by the representative vehicle to surrounding links while the representative vehicle expands its signal range. A simulation of the vehicle information transmission method with incorporation of the vehicle movement on a digital map of the actual area revealed that the transmission rate would vary according to the communication distance, and that the message deletion method and the link group method that we proposed would contribute more to increasing the transmission rate than simple flooding would.

Furthermore, to detect a possible traffic jam, we proposed a means to use vehicle information at certain timing and to assign a weight to each vehicle depending on where it is located on the link. Then we examined possible applications. Issues that must be resolved are the need to determine, using simulation, the appropriate range for transmission, the message-sending interval, and the range at which vehicle information is given.

## REFERENCES

[1] Nine Development Areas of ITS (Ministry of Land, Infrastructure, Transport and Tourism), http://www.mlit.go.jp/road/ITS/topindex/topindex_g02 _1.html

[2] Kaoru Tamura and Makoto Hirayama, Toward Realization of VICS – Vehicle Information and Communications System, IEEE Vehicle Navigation and Information Systems Conference (VNIS'93), pp.72–77 (1993).

[3] VICS/Navigation (Ministry of Land, Infrastructure, Transport and Tourism), http://www.mlit.go.jp/road/ITS/topindex/topindex_g03 _3.html

[4] Tetsuo Horimatsu, Standardization of Inter-Vehicle Communication, IPSJ Symposium Series 2007(2), pp.27-44 (2007).

[5] Sadayuki Tsugawa, Inter-Vehicle Communications and their Applications to Intelligent Vehicles: An Overview, Intelligent Vehicle Symposium, June 17-22 (2004).

[6] Tamer Nadeem, Sasan Dashtinezhad, and Chunyuan Liao, Traffic View: A Scalable Traffic Monitoring System, 2004 IEEE International Conference on Mobile Data Management (MDM), January 19-22 (2004).

[7] Jun Tsukamoto, Masashi Saito, Takaaki Umedu, Teruo Higashino, Inter-Vehicle Ad-hoc Protocol RMDP for Acquiring Preceding Traffic Information and its Performance, Transactions of Information Processing Society of Japan 47(4), 1248-1257 (2006).

[8] Kenta Omi, Takeshi Nishimura, Hisao Koizumi, Jun Sawamoto, A Real-time Traffic Flow Measurement Method based on Inter Vehicle Communication Considering the Link Group and the Loading Ratio of Inter Vehicle Communication Device, IPSJ SIG Technical Report 2007-ITS-28 pp.7-14 (2007).

[9] Kenta Omi, Takeshi Nishimura, Hidetoshi Kanbe, Hisao Koizumi, Jun Sawamoto, A Real-time Traffic Flow Measurement System based on Vehicle Information by Inter-Vehicle Communication, IPSJ SIG technical reports 2007(116), 9-16 (2007).

[10] Peter Davis, Satoko Itaya, Jun Hasegawa, Akio Hasegawa, Naoto Kadowaki, Akira Yamaguchi, Sadao Obana, Analysis of Characteristics of Flooding for Inter-Vehicle Communications, IPSJ SIG Technical Report, 2005-ITS-23, pp.99-104 (2005).

[11] S.-Y. Ni, Y.-C. Tseng, Y.-S.Chen, and J.-P. Sheu, The broadcast storm problem in a mobile ad hoc network, Mobicom'99, pp.151-162 (1999).

[12] Y.-B. Ko and N. H. Vaidya. Location-aided routing in mobile ad hoc networks, Mobicom, pp.66-75 (1998).

[13] Sugiyama Yuki, Physics of Traffic Flow, Journal of Japan Society of Fluid Mechanics 22(2), 95-108 (2003).

[14] ESRI Shapefile Technical Description http://www.esri.com/library/whitepapers/pdfs/shapefile .pdf

# Session 3: Information System
## (Chairs Hidekazu Shiozawa and Naoki Nakamura)

# Remote Kenken: An Exertainment Support System using Hopping

Hirotaka Yamashita*, Junko Itou** , and Jun Munemori**

*Graduate School of Systems Engineering, Wakayama University, Japan
**Faculty of Systems Engineering, Wakayama University, Japan
{s105056, itou, munemori}@sys.wakayama-u.ac.jp

***Abstract*** - Games that use sensors and the physical body have become widespread. However, existing games have a problem that the playing position is fixed. In this study, we propose a system, "remote kenken," by which one can really move their body, and can exercise. We have overcome the limitation on body movement by arranging sensors in the same way as "kenken". In addition, a moderate exercise load was realized by repeating the game five times. We performed a comparison between the proposed system and a kenken game, in which a subject jumped over the same place. The proposed system was evaluated highly in the sense of reality.

***Keywords***: pressure sensor, exercise, game, entertainment, kenken

## 1 INTRODUCTION

TV games are one of the forms of entertainments that can easily be enjoyed within the family. Most of them mainly use a controller, but there are few sensor operating systems. There are some games for exercise and entertainment, which we call "exertainment". Physical exercise is recommended for senior citizens [1]. It is expected to become one of the future exercise methods. However, one's position is fixed in these games, and there is the problem that movement is limited. In other words, it is only a simulation of exercise.

We have proposed an exertainment support system, "remote kenken," which overcomes this limitation on movement. We simulate kenken by the proposed system using pressure sensors.

## 2 EXISTING SYSTEMS

Wiifit [2] is a popular software developed for the game console "Wii," released by Nintendo. Wiifit is a game aimed at doing exercise within the family. It is attached to a balance Wii board as a peripheral device. A balance Wii board is shown in Figure 1. Plural sensors are put on this balance Wii board and can sense careful movement such as weight movement.



Figure 1: Balance Wii board.

Family trainer [3] is software developed for "Wii," and is designed for exercise. A mat controller (shown in Figure 2) is attached as a peripheral device. It is simple structure. This system does not use a special sensor like a balance Wii board. "Kenken step" is included to perform hopping on the spot. Like Wiifit, Family trainer is limited to exercise on the mat controller.

The differences between remote kenken and kenken step are shown in Table 1.



Figure 2: Mat controller.

Table 1: Differences between remote kenken and kenken step

|  | Remote kenken | Kenken step |
|---|---|---|
| Moves | Jumps in the same way as in reality. | Jumps only on the spot. |
| Score | Step precisely on pressure sensors. + Playing time. | There is no indication of the score. Time should be fast. The player must jump as displayed on a screen. |

# 3   REMOTE KENKEN

## 3.1   Design policy

This system is performed in the same way as conventional kenken. We raise the sense of reality and aim at exercise. Therefore we use pressure sensors. We get the analog value of the position of a foot by pressure sensors. The system aims at exercise and entertainment.

## 3.2   Development environment

The system was developed by Visual C#. The program is about 1,800 lines of a client application, and about 600 lines in a server application. The circuit of the pressure sensor uses a PIC. 16F873-20/SP was used for processing data from a pressure sensor. The program is about 60 lines. The pressure sensor uses FlexiForce [4].

## 3.3   System constitution

This system consists of a PC (server and client) and pressure sensors (max 8). The circuit of the pressure sensor is shown in Figure 3, and a pressure sensor with acrylic board in Figure 4. Eight pressure sensors in a board are shown in Figure 5. The client manages the score. Two players are enabled by using a server application. The total constitution of this system is shown in Figure 6.



Figure 3: The circuit of the pressure sensor.



Figure 4: A sensor with acrylic board.



Figure 5: The pressure sensors on the floor.
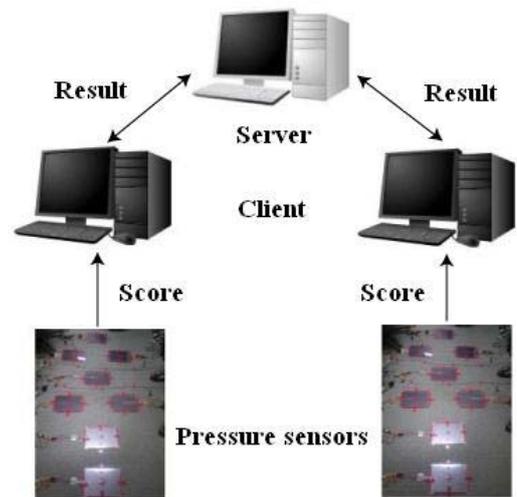


Figure 6: Total constitution of the system.

## 3.4 Calculation method of the score

As for the score, it is calculated by data from sensors. When a player steps on the center of the pressure sensor, two points are awarded. When a player steps any place other than the center of the pressure sensor, one point is awarded.

The normal mode is 32 points by stepping forward on the sensors. Furthermore, the score by playing time is expressed at 8 points. Total score is 40 points in maximum (8 points + 32 points). The five round trips mode is 160 points by stepping forward on the sensors. A score by playing time is expressed at 40 points out of a total of 200 points, which is a perfect score. The ratio of the score is 4:1. We put weight on the accuracy of steps.

Remote kenken usually performs this with eight pressure sensors connected.

The score of the player becomes 0 when a player didn't step on any pressure sensors. It is imperative that the player step on the first and the last sensor, otherwise the game cannot proceed.

As an example of the normal mode, a player got 20 points with the pressure sensor. When the playing time was 10 seconds, 2 points were added. The total score becomes 22 points in this.

## 3.5 Sound effects at the time of the play

Seven kinds of sound effects are used in the system. This is because a player can grasp present status even without watching the screen. When a player gets a point, we use two kinds of sound effects, so the player comes to understand whether he placed his foot on the center or not.

## 4 EXPERIMENTS AND RESULTS

### 4.1 Experiment summary

We carried out the experiment into whether it was useful as exercise. Subjects measured their pulse before and after playing the game. We assumed two players in the game simultaneously and compared the score of the game with partner.

### 4.2 Experiment environment

Subjects were eight persons at Wakayama University. A scene of the experiment is shown in Figure 7.

A subject pushes the standby button after having been connected to the server. The game starts. Subjects performed five round trips. Subjects step on the pressure sensor. Subjects repeated "kenken" five times from the starting point.


Figure 7: A scene of the experiment.

### 4.3 Results of application experiment

The results of the experiment are shown below.

The scores and pulse rates of the subjects are shown in Table 2. The average time of playing is 42.5 seconds. (Reference: The average of "kenken step" was 33.9 seconds). The average score was 129.6 / 200. The pulse rate rose 20.6 per one minute after the game. The average of "exercise strength" became 17.6%. "Exercise strength" is shown below.

Exercise strength (%) = (The pulse rate while exercising - The pulse rate at the time of rest) / (The best pulse rate - The pulse rate at the time of rest) $\times$ 100

Table 2: The score and pulse rates of the subjects

| Playing time (seconds) | Score | Pulse rate | per minute |
|---|---|---|---|
| | | before | after |
| 33.2 | 128 | 99 | 115 |
| 38.1 | 131 | 82 | 108 |
| 60.7 | 134 | 65 | 82 |
| 42.4 | 112 | 86 | 108 |
| 44.5 | 122 | 80 | 95 |
| 37.8 | 127 | 87 | 116 |
| 38.4 | 127 | 87 | 112 |
| 44.9 | 156 | 77 | 92 |

The results of the questionnaire are shown in Table 3 (1: very poor, 3: neither good nor poor 5: very good).

The description part of the questionnaire is as follows.

(1) Is the score reasonable?
・The ratio of weight for playing time and score from sensor is reasonable.

・If I can't step on the sensor, I think zero score is reasonable.

・The parameters of playing time are reasonable.

(2) Concerning others opinion and impression.

・I enjoyed it. There should be several variations about the kenken.

・I was worried about the delay of the sound.

・It is hard to understand the lap now.

・There should be comments from the system depending on your score. By getting a point, it should say "do your best!"

Table 3: Questionnaire results of experiments

| Questions | Evaluation average |
|---|---|
| As for playing time, was it reasonable? | 4.5 |
| Were you able to understand the rules? | 4.9 |
| Were you able to step well? | 3.4 |
| Did you understand whether you were able to place your foot by sound? | 4.4 |
| Was it like real hopping (kenken)? | 4.5 |
| Do you feel it was "exercise"? | 4.4 |
| Were you motivated more by a result of victory or defeat? | 4.5 |
| Do you think the setting of the score is reasonable? | 4.1 |
| Were you interested | 4.6 |

## 4.4 Consideration of the experiment

The results of the questionnaire on "remote kenken" and "kenken step" shown in Table 4. We shall next discuss the results of experiments based on Tables 3 and 4.

Table 4: Comparison with kenken step

| Questions | Remote kenken | Kenken step |
|---|---|---|
| As for playing time, was it reasonable? | 4.5 | 3.9 |
| Were you able to understand the rules? | 4.9 | 3.7 |
| Were you able to step well? | 3.4 | 2 |
| Was it like real hopping (kenken)? | 4.5 | 2.1 |
| Do you feel it was "exercise"? | 4.4 | 3.6 |
| Were you interested? | 4.6 | 3.9 |

(1) Concerning playing time and the understanding of the rules.

Average playing time for remote kenken took 42.5 seconds. The evaluation was 4.5. The length of playing time can be said to be reasonable. It took 33.9 seconds in average playing time for kenken step. Because the average of evaluation was 3.9, we thought that playing time of around 40 seconds is reasonable.

The average evaluation was 4.9 for understanding of the rules. That evaluation is higher than kenken step. Five round trips was easy to understand.

(2) Concerning sensors and sound effects.

The evaluation of "Did you understand whether you were able to place your foot by sound?" was high (4.4 / 5.0). But, pressure sensor sometimes sent incorrect data, and the players misunderstood their situation. A variety of sounds related to the lap is needed.

(3) On hopping sense and exercise.

The evaluation of the sense of reality was high (4.5 / 5.0).The evaluation of sense of reality was higher than kenken step (2.1 / 5.0).

The evaluation of "exercise" became 4.4. There were many people who felt it was "exercise". As for the pulse rate, an average of 20.6 per one minute increased. Remote kenken seems to be exercise.

(4) The score and victory or defeat indication.

Victory or defeat indication added motivation to the game. It was highly evaluated (4.5 / 5.0).

There were answers such as "the parameters of playing time were reasonable" for a description questionnaire. It was popular to get numerous points from pressure sensors.

(5) Concerning fun.

Evaluation of whether it was "interesting" was 4.6 which was higher than that of kenken step. There were opinions of "I enjoyed it. There should be several variations of kenken". By this experiment, we fixed the position of the pressure sensor. The sense of reality was kept in the system. If sound effects are added, the evaluation might go up.

## 5 CONCLUSION

In this paper, we proposed an exertainment support system named "remote kenken".

Remote kenken is a system in which players can jump around in the same way as real hopping by stepping on pressure sensors on the floor. We experimented 8 times (8 people) using the system.

The results of the application experiment are as follows.

(1) The sense of hopping seems to be highly evaluated in comparison with "kenken step," which simulates hopping.

(2) Around 40 seconds were judged to be reasonable in playing time. Because "kenken step" is thought to be reasonable at around 40 seconds, we thought such a small amount of time was reasonable for "kenken".

(3) After each experiment, pulse rates rose, so its value as exercise was accepted.

(4) Displaying victory or defeat by the score was evaluated as fun.

Sense of reality was high in this system. In the future, we will improve the system for exertainment and perform experiments between remote players.

## REFERENCES

[1] K.Yamada, M.Uematsu, A change of a physical balance function by a game of a healthy elderly, Healthy recreation study memoirs, vol.4, pp. 35-238, (2007).

[2] Wii Fit, http://www.nintendo.co.jp/wii/rfnj/.

[3] Family trainer, http://familytrainer.jp/.

[4] FlexiForce, http://www.nitta.co.jp/product/mechasen/sensor/flexi_summary.html

# A Proposal of an Adaptive P2P Database Retrieval System Using a Hierarchically Categorized DHT Structure

Takuya Sasaki[*], Hideyuki Akasaki[*], 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
g031e078@edu.soft.iwate-pu.ac.jp, sawamoto@iwate-pu.ac.jp

*Abstract* –In recent years, there are various information sources in our surroundings in accordance with the development of ubiquitous equipments and information and communication technology. However, as for the management 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. In this paper, we propose an adaptive database search system GrowApp which unifies multiple heterogeneous DB's in ubiquitous Peer-to-Peer (P2P) network environment. In GrowApp, we apply the mobile agent technology in DB search mechanism and build a virtual DB environment integrating multiple heterogeneous DB's in P2P network system. We introduce hierarchical grouping structure into DHT in P2P and enhance conventional keyword search capability. In addition, we consider mechanism of automatic generation and reformation of grouping structure based on the users' retrieval records. Finally, we discuss implementation and evaluation of the system.

*Keywords*: P2P network, virtual database, mobile agent, database search, data mining.

## 1 INTRODUCTION

In recent years, as the ubiquitous environment spreads, various services are requested to be provided according to the place (location) and time, etc. for the user who acts while moving around with a mobile device. The free mobile computing environment with an ad hoc communication environment like wireless LAN and the short distance wireless telecommunications, etc. is coming to be achieved and removing the restraint on user's movement [1].

On the other hand, it is said that the technology that enables the discovery and integration of dynamic services is indispensable as the mechanism that provides ubiquitous services. It is requested to adjust to user's favor, and the change in the context including location, time and user's situation dynamically. The application of technologies of semantic Web and the Web service, etc. is thought as a mechanism that dynamic services corresponding to such ubiquitous environment are achieved [2]. However, the problem of the management of information sources remains[3].

In ubiquitous environment, there are various information sources in our surroundings in accordance with the development of ubiquitous equipments and information and communication technology. It becomes possible for data sources distributed in various places to tie to the data processing nodes along with the existence of the ubiquitous network environment. However, as for the management 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.

In this paper, we propose an adaptive database retrieval system GrowApp which unifies multiple heterogeneous DB's in ubiquitous Peer-to-Peer (P2P) network environment. In GrowApp, we apply the mobile agent technology in DB retrieval mechanism and build a virtual DB environment integrating multiple heterogeneous DB's in P2P network system.

We introduce hierarchical categorization (grouping) structure into distributed hush table (DHT) in P2P and enhance conventional keyword search capability [4, 5, 6, 7]. Individual data bases are integrated by using the virtual schema [9], and all the database searches are enabled with scanning each database by the agent. Integrated databases are viewed as one P2P node and registered to the DHT and managed. Moreover, retrieval by category is enabled by using a hierarchically categorized DHT structure [8].

In addition, we consider mechanism of automatic generation and reformation of grouping structure based on the users' retrieval records. In the 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 discuss implementation and evaluation of the system. In the present study, it is targeted to simulate and to evaluate whether the automatic grouping of the DHT is carried out in P2P environment. The Kademlia algorithm [6] is considered appropriate for large number of nodes and is used in this simulation.

## 2 GrowApp SYSTEM

### 2.1 Features

GrowApp is an intellectual search engine intended for the user who uses the mobile communication environment. It aims to retrieve information that exists in all places of the society.

GrowApp is not a centralized database management system. In recent years, the number of ubiquitous databases which are managed and maintained by the individual, the group, and the store, etc. is increasing because of the development of the network system. GrowApp aims to combine those existing databases and to offer profitable information for the user.

The flow of the system until the user, using the mobile communication terminal, retrieves information, and acquires it is as follows.

(1) The user accesses GrowApp.
(2) The user inputs keywords to the system for retrieval.
(3) The system that acquires the keywords, generates the retrieval agent who possesses user's context and DHT.
(4) The retrieval agent looks up DHT with the hashed values of the user's retrieval keywords.
(5) When the node with contents for the value of the retrieval keywords is found, the retrieval agent moves to the node.
(6) The retrieval keywords are thrown to the virtual database management agent at the node with contents.
(7) The virtual database management agent who receives the retrieval keywords executes query to the virtual database at the node.
(8) The acquired value, by the execution of the query, is returned to the retrieval agent.
(9) The retrieval agent who acquires the value returns to the original user's position, and presents the acquired information to the user.

## 2.2 Structure

GrowApp is composed of four major elements of virtual database, mobile agents, the P2P network, and the DTH.

## 2.2.1 Virtual Database [9]

Various kinds of databases exist in the society. Because the management technique is different, they cannot adopt the same query method. To retrieve, and to update them from the application side, regardless of the kind of the database, the virtual database schema is set up between the user and databases. It converts the search queries and the virtual database schema doesn't ask the kind of the database. Figure 1 shows that the user accesses various databases without considering heterogeneity of each database by using the virtual database schema.

Various databases are assumed to be a retrieval object in GrowApp. It becomes a procedure of the database management agent's to execute the query to the virtual database schema and then the virtual database schema retrieves databases and acquires information.
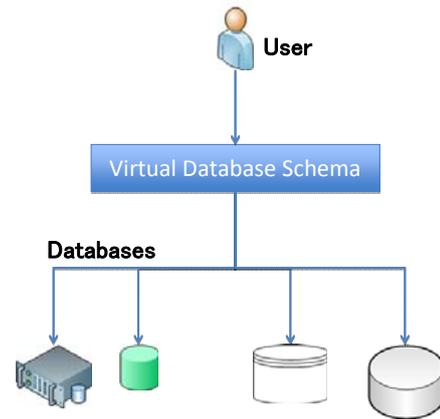


Figure 1: Structure of virtual database.

## 2.2.2 Mobile agent

A virtual database as stated in section 2.2.1 could include all the databases in society under its schema, forming a huge virtual database. To view all the databases under unified schema is ideal. However, we can easily see a problem of communication bottleneck between multiple users and the virtual database schema and it is hard to consider a unique virtual database schema.

In this paper, the virtualization is done by using the virtual database schema for a certain number of databases. The database management agent is arranged in each virtual database where those virtualizations were done.

In addition, when the user uses the system and the retrieval keyword is turned on, the system generates the retrieval agent who has user's context. It is assumed that user's context is the one to show the situation that relates to the user at the point of the retrieval. A past retrieval history can be easily reflected in a present retrieval by adding a past retrieval history in addition to spatial information and temporal information to the context.

When the retrieval agent retrieves data as shown in Figure 2, query is thrown out to the database management agent. The database management agent, who receives query, retrieves data from the heterogeneous databases by executing query for the virtual database schema.
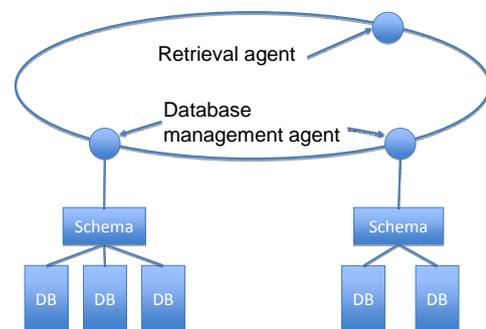


Figure 2: Retrieval agent and database management agent

## 2.2.3 P2P network and DHT

When the database is brought together in one place, the bottleneck between databases and users is caused. To solve

his problem, when the user retrieves data, a retrieval agent is generated, and information is acquired according to the P2P communication between agents.

Ad hoc participation of the nodes to the network can be expected. Then, assuming frequent arrival and departure of participating nodes, the DHT that uses Kademlia algorithm [6] is introduced, and the P2P network is constructed. As for the Kademlia algorithm, it is evaluated as comparatively easy to implement because special update work is unnecessary under such frequent updates condition.

## 2.3 Functions

GrowApp has the following functions as an intelligent search engine that considers the mobile computing environment intended for a lot of heterogeneous databases.

### 2.3.1 Current user location

The user is expected to participate in an ad hoc system in the mobile computing environment. Then, user's spatial information is acquired, and the spatial information is treated as user's context. The user's spatial context is used to understand the user's situation and to put the weight to most suitable information.

When the user accesses the system, because the GPS function is installed in today's cellular phone, user's GPS information can be acquired. This information is acquired, and maintained by the generated retrieval agent of each user.

### 2.3.2 Grouping of contents in DHT

This system constructs the P2P network. When the node retrieves contents, the DHT is used. Retrieved contents are downloaded by tracing down the node that maintains contents from the query that the user executed referring to the DHT. 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.

Then, the DHT is hierarchically categorized, and contents are grouped. Flexibility can be introduced into the user's retrieval by providing information with high possibility to be retrieved at the same time to the search keyword by this contents grouping. It is thought that it can be supplemented to lack of a flexible retrieval function that is the weak point of the DHT.

### 2.3.3 Dynamic computation of category by data mining

When the user retrieves many of the databases that exist in the society, a large amount of information will be presented to the user. It is very difficult to discover profitable information for the user [10].

Then, the category is measured by data mining. The retrieval agent maintains a past retrieval history of the user besides user's positional information. When the user uses the

system and retrieves data, the preference of user's retrieval history and acquired information is measured. For the measurement method, the system puts higher priority for contents the user downloaded in the past retrieval. It becomes possible to present information that is thought to be a reflection of the result of the measurement and effective to search preferred data from among a lot of retrieval results.

## 3 AUTOMATIC GROUPING IN DHT

### 3.1 Outline

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 retrieval results frequently. Then, we propose an automatic grouping function of the DHT. It becomes possible to offer contents grouping and renewal of DHT in real time by this function.

In this paper, the Kademlia algorithm 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. When a new category, in addition to participation and the secession of the node, was created and an old category is deleted the hush table is renewed. 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 situation.

It is possible to view those retrieval contents as one group when a certain user uses the system, and user retrieves not only one 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 category. However, when a flow of time and a new kind of item appear, the item that comes off as a retrieval target is not permanently left as a member of category but the category is updated in accordance with the retrieval frequency etc. It becomes possible to make the retrieval efficient and suitable for the preference that the user is always requesting.

### 3.2 Example of group creation and deletion (example of PC parts)

An example flow of the automatic creation of a new group that occurs when the user actually retrieves and contents are downloaded, and deletion of a group when it becomes not used, is described as follows.

(1) Contents registered in the DHT do not belong to a group in the initial state. Contents are shown in Figure 3. "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 retrieved often at the same time. Grouped contents are shown in Figure 4.

(4) DVD drive appears. The DVD drive is a content named "DVR-216DBK" here. The state is shown in Figure 5.

(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. Figure 6 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 Figure 7. However, only the group of the CD drive is deleted, and CD drive itself is possible to be retrieved and downloaded as a content.



Figure 3: Initial state without a group



Figure 4: A group is generated



Figure 5: Appearance of DVD drive
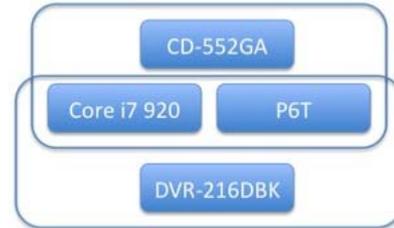


Figure 6: New contents group is generated



Figure 7: The group that contains the CD drive is deleted

## 4 SYSTEM IMPLEMENTATION

The system consists of the User Interface part, mobile agent, P2P network, and the virtual database. Programming language Python is used for implementation. The function of contents grouping in DHT is implemented on the P2P simulator with Kademlia algorithm, and the usability is experimented and evaluated. Then, we continue to build the virtual database schema, the retrieval process, the database management agent and the user interface of GrowApp.

User Interface has the function to display information to the user and the form that input the retrieval keywords. This system uses two kinds of mobile agents. One is a retrieval agent who receives user's retrieval keyword and starts retrieval of contents information. And, the other is a database management agent who receives contents information in the virtual database and query from the retrieval agent. For the virtual database, multiple databases are virtualized by using the virtual database schema prototyped by the project [9].

## 5 EVALUATION

Here, we measure the communication traffic to estimate the degree of burden to the user to see whether the automatic grouping of contents in the DHT is useful or not. And at the same time, retrieval performance is evaluated quantitatively.

### 5.1 Outline of the evaluation

Outline of the evaluation is shown as follows. Number of PCs (install about 200 nodes per PC) is increased and the following parameters are measured.

・ Measurement of traffic volume of information
・ Investigation of retrieval performance (precision, recall)
・ Response time of retrieval

In the evaluation, about 200 nodes are generated per one PC, and P2P network is constructed. And the measurement of traffic is carried out and retrieval performance with the DHT is verified.

## 5.2 System structure

Figure 8 shows the system structure of the evaluation. In the evaluation, in order to evaluate the accuracy of the DHT with automatic grouping and the communication traffic, we generate more than 200 nodes and construct simulated P2P network. Whether contents in the DHT are added accurately as a contents group based on user's retrieval keyword, and how communication traffic changes compared with the client-server model are verified.
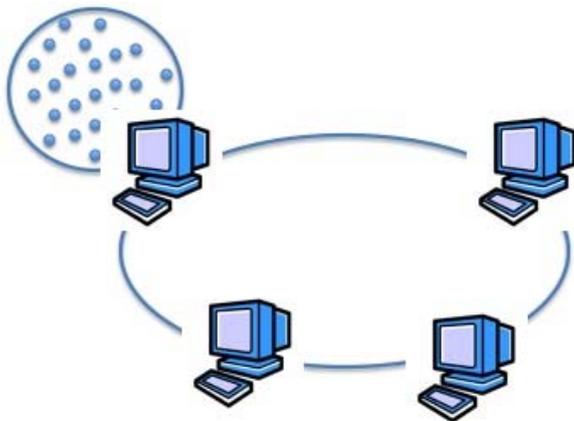


Figure 8: System structure of the experiment

## 5.3 Software configuration

Figure 9 shows the software configuration. The P2P simulator is started on each terminal. As use of various mobile communication terminals are assumed, both Windows and Linux are used for OS of experiment. Put the hush operation on the retrieval keyword from the user and refer to the DHT for the P2P simulator. A mobile agent moves to the specified node position, and the mobile agent communicates with the database management agent, and demands and acquires contents information that the user is requesting. Contents information is displayed to the user at the user's interface.
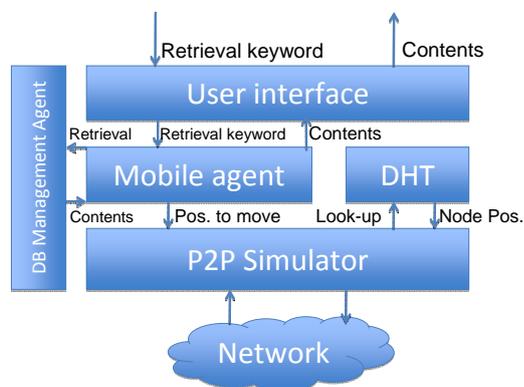


Figure 9: Software configuration

## 5.4 Evaluation

Items of the evaluation corresponding to the content of the experiment are as follows. The client-server model that stores the same volume of information as the experiment system in a database is used for comparison.

- Comparison of communication traffic when the experiment system and client-server model carry out the same retrieval operation.
- Accuracy evaluation of contents that are retrieved with keyword search.
- Confirmation of automatic grouping of contents in the DHT when multiple contents are retrieved simultaneously.
- Comparison of response time when contents are retrieved by this system and client-server model.
- Evaluation of the system under the condition of increasing number of participating nodes.

Based on the results of the experiment, the contents grouping in DHT is evaluated whether a useful measure or not. Evaluation criteria are communication traffic for DHT construction, retrieval, update, comparison with client-server model, retrieval performance for the user request. As a result of the actual experiment, the function of contents grouping of the DHT will be improved in the next step. Then, the implementation as a prime function of GrowApp will be examined.

## 6   CONCLUSION

In this paper, we proposed an adaptive and intelligent information retrieval system GrowApp targeting existing databases in the ubiquitous network society. We intend to implement and evaluate the GrowApp's  primary function of grouping of contents in DHT. Based on the evaluation, improving the function of grouping of contents, the interface part with the virtual database schema, mobile agent, and user interface part will be integrated, and it is scheduled to go finally the system design of the entire GrowApp. In addition, the verification of the usability is scheduled to be covered in consideration of a concrete application system of the proposed GrowApp, and experimenting with the application.

## ACKNOWLEDGMENT

## REFERENCES

[1] D. Chakraborty et al., "Toward Distributed Service Discovery in Pervasive Computing Environments," IEEE Trans. Mobile Computing, vol. 5, no. 2, pp. 97–112 (2006).

[2] Hiroyuki OGURA, Saeko MURAKAMI, Hiroyuki SATO, Tomihiko KOJIMA, Noboru SHIMIZU, Itaru HOSOMI, Application Systems on Semantic Web, Journal of Information Processing Society of Japan 43(7)  pp.742-750 (2002).

[3]Ryan Huebsch, Brent Chun, Joseph M. Hellerstein, Boon Thau Loo, Petros Maniatis, Timothy Roscoe, Scott Shenker, Ion Stoica, and Aydan R. Yumerefendi. The architecture of pier: An internet-scale query processor. In The Second Biennial Conference on Innovative Data Systems Research (CIDR), Asilomar, CA, January (2005).

[4]Ben Y. Zhao, John Kubiatowicz, and Anthony Joseph. Tapestry: an infrastructure for fault-tolerant wide-area location and routing. Technical Report UCB/CSD-01-1141, U.C. Berkeley, April (2001).

[5] Ion Stoica, Robert Morris, David Karger, M. Frans Kaashoek, and Hari Balakrish-nan. Chord: A scalable peer-to-peer lookup service for internet applications. In Proceedings of the ACM SIGCOMM '01 Conference, San Diego, California, August (2001).

[6]Petar Maymounkov, David Mazi`eres, 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), March (2002).

[7] HIROKI KUZUNO, YUSHI NAKAI, ATSUMU WATANABE, TAKUYA KAWAHARA, TAKASHI KATOH, BISTA BHED BAHADUR, TOYOO TAKATA, A Proposal of Distributed Internet Monitoring System Using Mobile Agent, Transactions of Information Processing Society of Japan 47(5) pp.1393-1405 (2006).

[8] Yi WAN, Takuya ASAKA and Tatsuro TAKAHASHI, A Hybrid P2P Overlay Network for Non-strictly Hierarchically Categorized Content, IEICE Transactions on Communications E91-B(11):3608-3616 (2008).

[9] Y. Watanabe, K. Shoubu, H. Miita, Y. Wada, J. Sawamoto, T. Katoh, Virtualization technology for heterogeneous databases, FIT2009 (2009).

[10]K. Mori, S. Kurabayashi, N. Ishibashi, Y. Kiyoki, An active in formation delivery method with dynamic computation of user's information in mobile computing environments, IEICE, DEWS2004, March (2004).

# Proposal and Practice of Exchange Learning through Quiz Creation and Peer Review

Masanori Takagi[*], Masahiro Mochizuki[**], Masamitsu Mochizuki[***], and Yoshimi Teshigawara[**]

[*] Faculty of Engineering, Soka University, Japan
[**] Graduate School of Engineering, Soka University, Japan
[***] Faculty of Business Administration, Soka University, Japan
{mtakagi, e09m5239, mochi, teshiga}@soka.ac.jp

***Abstract*** - In this paper, we propose a method of learning that includes interactively reviewing and taking quizzes created by learners of different classes. First, we describe a brief outline of the CollabTest system that we have developed and evaluated. Next, we describe the proposed method of learning exchange between classes, including the details of quiz creation, mutual review, and taking quizzes using the CollabTest system. We discuss the expected effects of the proposed method on learners. In addition, we implement the CollabTest system so that learners can study using the proposed learning method. A trial study was conducted in the semester beginning in April 2009, and as a result of these experiments, we determined the requirements and effectiveness of learning through collaboratively creating and taking quizzes created by learners in different classes.

***Keywords***: Exchange Learning, Quiz Creation, CSCL, Peer Review, Online Test.

## 1   INTRODUCTION

We studied a web-based learning support system named "CollabTest" that enables learners to acquire knowledge by creating quizzes and sharing them with peers [1, 2]. The learning procedure when using CollabTest is as follows.

I.   Learners in a class create multiple-choice quizzes with explanations of the content.
II.  They review the quizzes collaboratively in a group.
III. They submit the quizzes to their teacher.
IV.  Learners take student- or teacher-created quizzes to confirm their comprehension levels.

We have developed the CollabTest system to enable students to perform the above-mentioned tasks on the Web. Moreover, we have continually used this system since 2002 at schools of various levels, including a university, high school, elementary school, and vocational training school. In 2003, we developed a point system in which students can earn points by performing tasks such as submitting a quiz or posting a comment. Furthermore, we allowed learners to compete among individuals or among groups via a point-ranking function. We have used this system in a total of 104 classes over 6 years. As a result, 5083 learners have used the system, 14456 quizzes have been created by learners, and 40652 comments have been posted by learners. These results show that CollabTest can help resolve issues such as the lack of e-Learning materials and the lack of student-student and teacher-student interactions.

Exchange learning, which involves exchanging information between regional schools via the Internet and a video conference system, has been actively studied at junior high and elementary schools in recent years [3-8]. However, practical studies of exchange learning through quiz creation and peer review have not yet been reported. In addition, practical studies of question-posing and peer-assessment via web-based systems that facilitate learning between classes have not been reported in related studies on learning systems. Exchange learning using a system such as CollabTest requires neither preparation of learning materials nor scheduling, as is the case when using a video conference system, as the environment of CollabTest is asynchronous. Thus, we can expect that exchange learning between classes can be conducted more easily than other methods of exchange learning. Moreover, we expect to observe considerable effects due to communicating between classes in addition to those of CollabTest [1, 2].

For these reasons, CollabTest is thought to be an effective system for exchange learning; however, we have used it only in a closed environment, for example, at a particular university. We have considered expanding the usage environment of CollabTest to include an open environment such as the World Wide Web where various learners could participate. With this in mind we apply CollabTest to exchange learning in order to expand its versatility and create new value. In this paper, we consider learning methods for exchange learning using CollabTest and their effects. We then clarify the functionality that is necessary for exchange learning and report the results of using one learning methods that we considered.

## 2   COLLABTEST

We have developed CollabTest as a Web application using JSP (Java Server Pages) and Java servlets. The system provides a group management function that can divide students into groups, a quiz entry function that is used by not only teachers but also learners, and an online test function that can deliver tests based on the quizzes created by the students and teachers. A noteworthy feature of this system is the collaborative environment for quiz creation. The environment provides a peer-review function that shares a learner's quiz with their group and allows group members to interactively assess the quiz. As shown in Figure 1, learners can assess quizzes created by group members according to review items displayed in the system. In addition to these functions, it provides a point system in
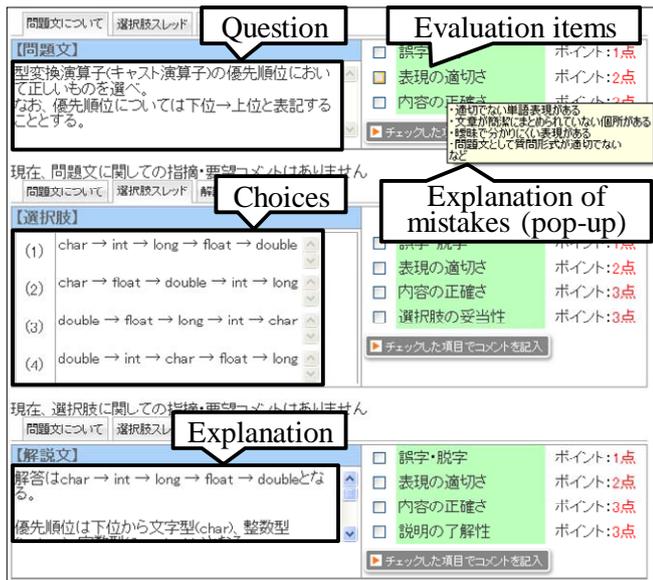
Figure 1: Review page for a quiz.

which learners accumulate points through the use of the system, as well as a point-ranking function that allows competition between individual learners or groups.

# 3 EXCHANGE LEARNING USING COLLABTEST

## 3.1 CollabTest and Exchange Learning

Many case studies on exchange learning have been reported in recent years. Although the learning methods and contents of learning reported in these case studies were different, common points were reported regarding the information exchanged between classes. Based on these common points, exchange learning between classes is defined here as a learning method that involves sharing the knowledge of learners in different classes and exchanging comments or feedback on their respective knowledge. Since the classes are different, the learners gain access to different teachers, schools, cultures, and grade levels, as well as the different specialties, knowledge, and experiences of their peers. In CollabTest, sharing knowledge between classes occurs when taking quizzes and the exchange of information occurs when posting of comments on quizzes. In the following, learning is classified into two steps: evaluating quizzes interactively and taking the quizzes. Then, an example of an implementation of this learning method and the expected effects of the exchange using this system are described.

## 3.2 Exchange Method

### (1) Learning exchange through Peer review

In one peer-review activity of CollabTest, a learner shares his/her quizzes with members of the same group. If the learners share quizzes with learners in a different class, they can share not only knowledge but also exchange comments on their quizzes.
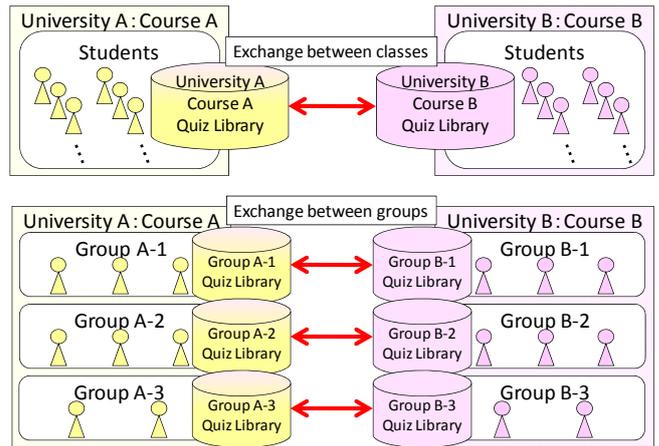


Figure 2: Methods for combining learners.
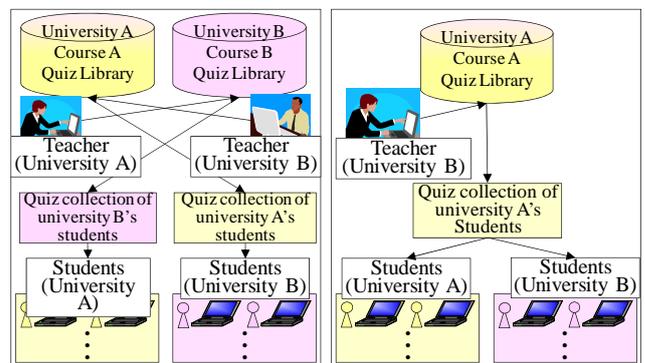


Figure 3: Method for creating a test.

As shown in Figure 2, we considered two methods of learning exchange through the peer review. The first method is an exchange between classes. The second method is an exchange between groups. In the case of reviewing interactively with other classes, learners can browse all quizzes created in the other class. As a result, a learner who is interested in the quizzes created in the other class has considerable opportunities for learning exchange. However, there is a possibility that an unmotivated learner who is not interested in the quizzes will not participate fully, especially if there are a large number of quizzes and the learner does not understand what should be done.

In the case of reviewing interactively between groups, a learner who is not interested in the quizzes can likely begin to learn more easily since the number of quizzes that they can browse is limited, thus making the material more accessible. On the other hand, a motivated learner will likely feel that the material is not sufficiently challenging. For these reasons, it is necessary to develop a function that enables learners to choose whether to learn between classes or groups depending on the number of learners, their motivation, and the particular circumstances of the classes.

### (2) Exchange through Taking Online Tests

In online tests, learners take quizzes created by peers in the same class, quizzes created by past classes, and quizzes created by the teacher. Learners can post comments such as their questions about and impression of each quiz using the online test function of our system. If the learner takes an online quiz created a different class, the learners can

Table 1: Examples of differing class attributes.

| No | Difference of class attribute ( -: same, ✓: different ) | | | | | Example |
|---|---|---|---|---|---|---|
| | Learners | Teacher | School | Grade or amount of knowledge | Specialty | |
| 1 | ✓ | - | - | - | - | • Exchange between classes that have the same course name and the same teacher at the same university. |
| 2 | ✓ | ✓ | - | - | - | • Exchange between classes that have the same course name and different teachers at the same university. |
| 3 | ✓ | ✓ | ✓ | - | - | • Exchange between classes at different universities that have the same or similar course names. |
| 4 | ✓ | ✓ | - | ✓ | - | • Exchange between fifth graders and sixth graders. |
| 5 | ✓ | ✓ | ✓ | ✓ | - | • Exchange between elementary school students and junior high school students.<br>• Exchange between junior high school students and high school students.<br>• Exchange between high school students and undergraduate students.<br>• Exchange between undergraduate students and graduate students. |
| 6 | ✓ | ✓ | - or ✓ | ✓ | ✓ | • Exchange between economics class for students specializing in economics and economics class for students specializing in other fields. |

mutually share knowledge and have the opportunity to post comments on the quiz after its completion.

As shown in Figure 3, we considered two methods of using quizzes created by learners in different classes. In the first method (left of Figure 3), the teacher composes the test by selecting questions from quizzes created by learners. In the second method (right of Figure 3), the teacher composes a test using quizzes created in his/her class and shares it with the other class. In the first method, the teacher can use the quizzes that he/she wants to use, although this requires a commitment of time from the teacher. In the second method, the teacher can administer the test without using a large amount of time for test preparation. However, a teacher should be able to confirm and edit the test contents before administering it to the learners as there is the possibility of including quizzes at an inappropriate level or containing inappropriate content for the class.

## 3.3 Expected Effects

In exchange learning using CollabTest, quizzes created by learners on the course content are used as learning materials. For this reason, different classes with similar course content can learn in collaboration. Moreover, a video conference system has been used to exchange comments in the related study on exchange learning. However, for this system, learners use a BBS (Bulletin Board System) in order to exchange comments, thereby it is not necessary to adjust the course schedules of the classes and the learning exchange can be implemented easily.

The expected effects of learning interactively between classes using CollabTest will be affected by the attributes of the classes participating in the learning exchange. These attributes include the teacher, school, and grade level, as well as the specialties, knowledge, and experiences of the learners. Table 1 shows an example of differing class attributes. The common effects of using CollabTest are expected to be as follows.
- The learners will participate in a class more actively as a result of their awareness of learners in the other class.
- The motivation for the class will improve as a result of taking quizzes created by the other class and receiving comments from learners of the other class.
- Communication skill will improve as a result of exchanging information with learners with whom they are not acquainted.

As class attributes become increasingly different, these effects are expected to become larger. Moreover, we expect that the more comments the learners exchange, the larger these effects will become. We expect the following effects when there is a difference in the grade level or amount of knowledge between classes participating in the exchange learning (Table 1, No. 4–6).
- The learners in the lower grades will be able to receive more advanced and specialized comments and suggestions.
- The learners in the higher grades or with the higher amount of knowledge will be able to improve their explanatory and descriptive skills since they must post easy-to-understand comments for learners with incomplete understanding of the learning contents.

## 4 FUNCTION ENHANCEMENT

It was necessary to develop new functions to enable mutual learning between classes using CollabTest. We considered the following to be necessary functions.
- Registering courses to initiate the exchange learning.
- Reviewing quizzes interactively between classes.
- Delivering quizzes interactively between classes.

For this study, we developed the first two functions list above.

Figure 4: Course registration page.
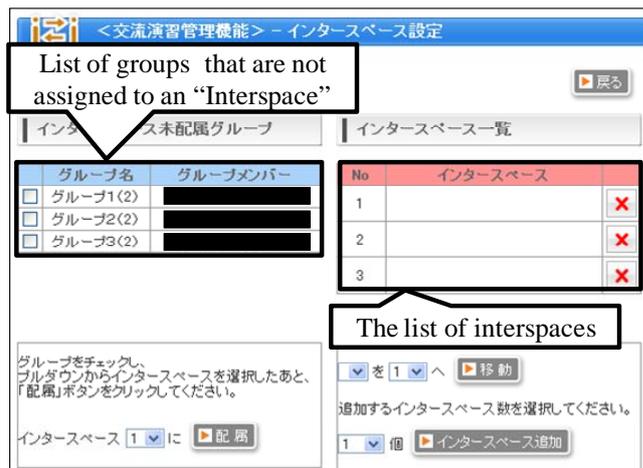


Figure 6: Interspace management function.
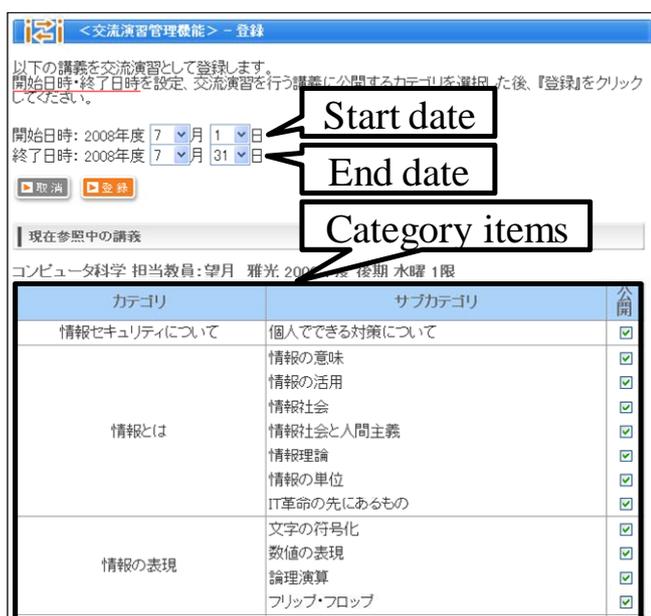


Figure 5: Initializing page for the exchange learning.



Figure 7: Function for displaying quizzes created by other class.

## 4.1 Administrative Functions for Exchange Learning

The administrative functions for the exchange learning system are provided by an administrator or teacher mode. Figure 4 shows the user interface of the function to register courses. An administrator or teacher can access the function after logging on and accessing a course page. This page displays a list of the courses that start in the same semester as the accessed course. The list displays categories describing the course content and course information such as the course name and the teacher's name. The category items correspond to the scope and learning units of quizzes. Learners consult this list when they create a quiz [2]. In CollabTest, learners must select a category item registered by their teacher when they create a quiz. Consulting this list provides the administrator or teacher with an outline of each course.

When the teacher or administrator clicks the registration button displayed next to the course information, he/she can

access the initializing page where he/she can set up the start and end dates of the exchange learning. Figure 5 shows the user interface of the initializing page for exchange learning. On this page, after the teacher or administrator sets up the start and end date of the learning, he/she selects the category items that can be accessed mutually during peer review between classes from the above-mentioned category list. Learners learn collaboratively through the peer-review system using the quizzes registered in the selected category items.

## 4.2 Peer-review Function for Exchange Learning

We developed a function, referred to as the "interspace management function", in order to bring together learners in different classes. We developed the interspace as a space to assess quizzes between groups in different classes. Figure 6 shows the interspace management function provided in administrator or teacher mode. The left side of Figure 6

displays groups that are not assigned to the interspace among the groups that the teacher registered in the course. The right side of Figure 6 displays the groups that are being assigned to each interspace. We can assign more than one group to an interspace. Each teacher assigns groups from his/her course to any of the interspaces. If the teacher assigns all groups to same interspace, the learners can learn between classes as shown in Figure 2. When the teacher assigns each group to separate interspaces, the learners can access quizzes created by members of the group assigned to the same interspace and post comments on the quizzes. Figure 7 shows a function for displaying the list of quizzes created by other groups assigned to the same interspace.

# 5 EXPERIMENTS

## 5.1 Overview of Experiments

We conducted two experiments in the beginning in April 2009 in order to investigate the effects of exchange learning between classes using CollabTest. Experiment 1 was conducted for two class of the same course, Programming Practice 1, which was taught by two different teachers at Soka University. In addition, Experiment 2 was conducted for two similar courses, Introduction to Computer Networks, which were taught by two different teachers at Soka University and Aichi Prefectural University. Table 2 shows an overview of these courses.

Programming Practice 1 (referred to as Prog. 1 and Prog. 2) is a beginners' class for the C programming language, and Prog. 1 and Prog. 2 (see Table 2) used the same textbook. Experiment 1 corresponds to No. 2 shown in Table 1. Introduction of Computer Networks (referred to as Network 1 and Network 2) is a course on topics such as the history of the Internet, OSI reference models, and TCP/IP. In addition, Network 1 and Network 2 (see Table 2) used the same textbook. Experiment 2 corresponds to No. 3 shown in Table 1. All teachers taught class for 90 minutes once a week.

In the first half of the semester, each teacher used CollabTest only in their class as it has been used in the past. Then, CollabTest was used for exchange learning between classes in the second half of the semester. In Experiment 1, there were two opportunities for exchange through peer review and taking quizzes, as described in Section 3.2. In the exchange through peer review, we assigned all groups to one interspace, thereby all learners were able to access all quizzes created by the two classes. In the Experiment 2, we conducted the exchange only through taking online tests since the experimental period was short and there was a large difference in the number of students in each class.

## 5.2 Result of System Usage

Table 3 shows the results of exchange through peer review. Students taking Prog. 1 created 49 quizzes that were shared with students taking Prog. 2, and 61 quizzes were created by students taking Prog. 2, which were shared with students in Prog. 1. However, there were only two or three students who

posted comments about quizzes created by students of the other class. Moreover, 4 out of 7 comments posted in Prog.

Table 2: Course overviews.

| No | Course ID | University | The number of Students | Experimental Period |
|---|---|---|---|---|
| 1 | Prog. 1 | Soka University | 46 | 2 months |
| | Prog. 2 | Soka University | 49 | |
| 2 | Network 1 | Soka University | 102 | 2 weeks |
| | Network 2 | Aichi Prefectural University | 9 | |

Table 3: Results of exchange through peer review.

| Course ID | Quizzes (*1) | Comments | Students-1 (*2) | Students-2 (*3) |
|---|---|---|---|---|
| Porg. 1 | 49 | 7 | 3 | 21 |
| Prog. 2 | 61 | 33 | 2 | 4 |

*1: Number of shared quizzes
*2: Number of students that posted comments
*3: Number of students that received comments

Table 4: Results of exchange through taking online tests.

| Course ID | Quizzes (*1) | Students-3 (*2) | Frequency (*3) | | Comments |
|---|---|---|---|---|---|
| | | | Average | SD | |
| Prog-1 | 12 | 27 (58.7%) | 1.4 | 1.4 | 0 |
| Prog-2 | 9 | 25 (51.0%) | 1.6 | 2.0 | 0 |
| Network-1 | 7 | 61 (59.8%) | 1.4 | 1.3 | 4 |
| Network-2 | 10 | 5 (55.6%) | 1.2 | 1.2 | 0 |

*1: Number of quizzes included in the online test
*2: Number of students that took the test
*3: Frequency at which students took the online test

1 and 26 out of 33 comments posted in Prog. 2 were posted after the last class.

Table 4 shows the result for learning between classes through taking an online test. From this result we confirmed that more than half the students in each course had taken the quizzes created by learners in the other class. On the other hand, in Network 1, only four comments were posted after taking the online tests. In the other classes, there were no comments. These results show that the learners shared only their knowledge not their opinions in these experiments. This result may be due to not having sufficient time to take and assess the quizzes created by the other class during class time.

## 5.3 Questionnaire

**(1) Experiment 1**
Figures 8-10 show the results of a questionnaire given to students of Prog. 1 and Prog. 2. As shown in Figure 8, only 34% of students responded that they had been conscious of the progress of the exchange, such as the number of quizzes

created and comments posted by other class. In Experiment 1, there was little opportunity to be conscious of learners in



Q : You were conscious of the progress such as the number of quizzes and comments from the other class.

( valid responses: 74 )

- Very true
- Somewhat true
- Don't know
- Somewhat untrue
- Very untrue

12.2% 21.6% 32.4% 18.9% 14.9%
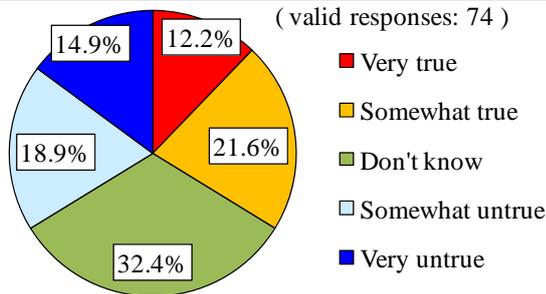
Figure 8: Results of questionnaire 1 (Experiment 1).



Q : How did you feel when you accessed or took quizzes created by students in the other class ?
Please select all appropriate choices.

1. It was interesting.
2. It was entertaining.
3. It inspired me.
4. It motivated me.
5. I wanted to answer more questions.
6. I did not feel anything in particular.
7. Other.
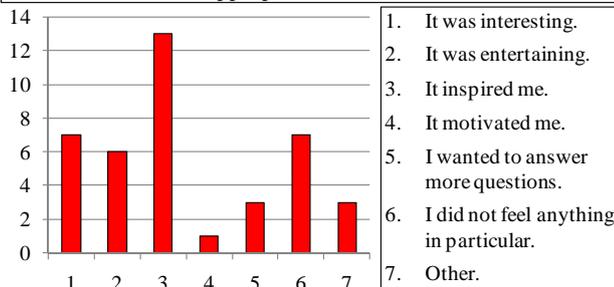
Figure 9: Results of questionnaire 2 (Experiment 1).



Q : I wanted to take quizzes created by students in a similar class at the other university.

( valid responses: 79 )

- Very true
- Somewhat true
- Don't know
- Somewhat untrue
- Very untrue

17.7% 35.4% 35.4% 6.3% 5.1%
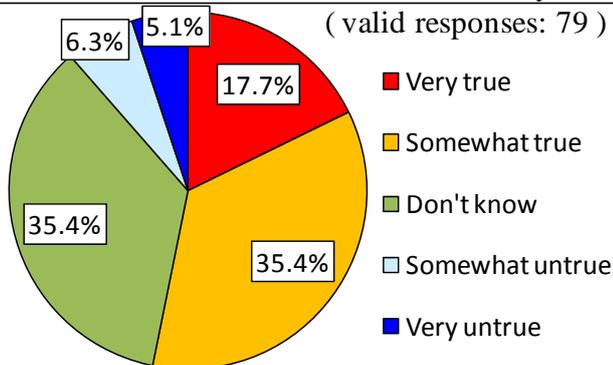
Figure 10: Results of questionnaire 3 (Experiment 1).

the other class since the few opinions were exchanged between the classes. It is possible that the results shown in Figure 8 are attributable to this. Figure 9 shows the results for a questionnaire given to students who had accessed or taken the quizzes created by students in the other class. As a result, 46.4% of students responded that they had been inspired by accessing or taking the quizzes. In the results of the questionnaire shown in Figure 10, more than half of the students responded that they wanted to take quizzes created by students in a similar class at another university. From these results, we confirmed the requirements for learning

through mutually taking quizzes that were shared between universities for students in a related class.



Q : How did you feel when you accessed or took quizzes created by students in the other class ?
Please select all appropriate choices.

1. It was interesting.
2. It was entertaining.
3. It inspired me.
4. It motivated me.
5. I wanted to answer more questions.
6. I did not feel anything in particular.
7. Other.
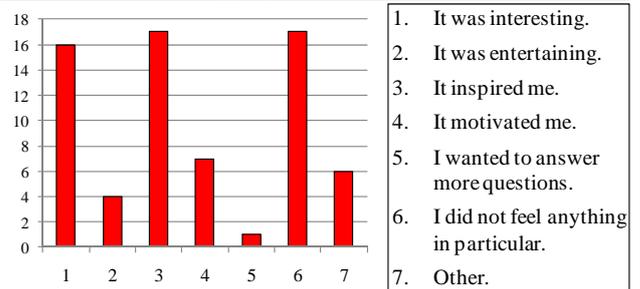
Figure 11: Questionnaire result-4 (Experiment-2).



Q : Taking quizzes created by students in a similar class at the other university was helpful for learning.

( valid responses: 58 )

- Very true
- Somewhat true
- Don't know
- Somewhat untrue
- Very untrue

20.7% 44.8% 20.7% 6.9% 6.9%
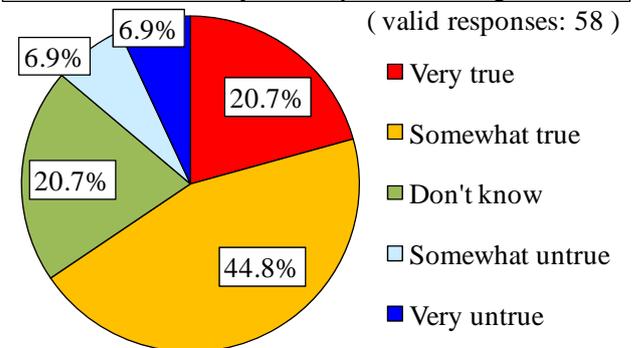
Figure 12: Results of questionnaire 5 (Experiment 2).

**(2) Experiment 2**

Figures 11-12 show the results of questionnaires given to students who took Network 1 and Network 2. The questionnaire in Figure 11 was given to students who had accessed or taken the quizzes created by students in the other class. Many students responded that they had been inspired by accessing or taking the quizzes. On the other hand, there were also many students who had responded that they had not felt anything in particular when they had accessed or took the quizzes. In the result of questionnaire shown in Figure 12, 65.5% of students responded that taking the quizzes had been helpful for learning. From these results, we confirmed the effectiveness of learning through mutually taking the quizzes created by students taking similar classes at different universities.

## 6 CONCLUSIONS

In this study, we proposed a method of learning through mutually assessing and taking quizzes created by learners in different classes. Moreover, we defined exchange learning between classes as follows.

- Learning through sharing knowledge with learners in a different class.
- Learning through mutually exchanging opinions about their knowledge.

Furthermore, we considered a method of exchange that included peer review and a method of exchange that included taking online tests. In addition, we evaluated the effects of these methods.

In experiments conducted at two universities, more than half of the students took quizzes created by students in the other class, and knowledge was shared between classes. However, there were only two or three students who exchanged comments between classes. In the results of questionnaires, we confirmed the requirements and effectiveness of learning through mutually taking quizzes created by learners in different classes, although only 34% of students responded that they had been conscious of the progress of other class. In the future, a comprehensive evaluation of the experimentation environment and manner of operation will be needed to ensure the active exchange of views between classes. For example, we will clarify the purpose or intention of the exchange learning in each class, and we will provide time to post comments and take the quizzes created by the other class during class time. In addition, we will clarify the effectiveness of the exchange learning using CollabTest through experiments involving exchange learning between classes with attributes other than No. 2 and No. 3 shown in Table 1.

## REFERENCES

[1] M. Takagi, M. Tanaka, and Y. Teshigawara, A Collaborative WBT System Enabling Students to Create Quizzes and to Review Them Interactively, Transactions of Information Processing Society of Japan, Vol.48, No.3, pp.1532-1545 (2007).

[2] M. Takagi, M. Tanaka, and Y. Teshigawara, Implementation and Evaluation on an Online Test System Enabling to Compete in Process of Creating Quizzes Collaboratively, Transactions of Japanese Society for Information and Systems in Education, Vol.24, No.1, pp.13-25 (2007).

[3] M. Kogawa, Y. Okazaki, K. Murai, T.Fujimoto, T. Yagyu, and M. Suzuki, A Study of International Exchange of Information about Children's Daily School Lives between Classrooms in Japan, the United States and Korea by Exchanging Videotapes, The Journal of school education，Vol.10, pp.153-164 (1998).

[4] T. Nagata, M. Suzuki, J. Nakahara, T. Nishimori, and T. Kasai, A New Teaching Practice for Prospective Teachers in a Family and Consumer Science Education Course : Various Kinds of Social Interactions in a CSCL Environment, Japan Journal of Educational Technology, Vol.27, pp.201-204 (2003).

[5] T. Nagata, M. Suzuki, J. Nakahara, T. Nishimori, and T. Kasai, The Effects of Interactions among Prospective Teachers who Studied Different Subjects in a CSCL Environment, Japan Journal of Educational Technology, Vol.28, pp.5-8, (2005).

[6] K Cho, Analysis of Intercultural Competency in the E-mail Exchange between Elementary School Students of Japan and Korea, Japan journal of educational technology, Vol.30, No.1, pp.59-67 (2006).

[7] A. Shikoda, K. Kazuo, K. Sugawara, S. Matsuzawa, T. Kawada, N. Kawada, I. Iguchi, T. Sato, and H. Sasaki, Practical Study and System Evaluation of High School and University Collaboration Work by Developing Embedded Educational Materials and Exchanging Teaching Assistants, Japan Journal of Educational Technology, Vol.32, No.2, pp.141-148 (2008).

[8] T. Inagaki, T. Horita, J. Takahashi, and H. Kurokami, Relations Between Practices and Communication Tools for Inter-School Collaborative Learning, Transactions of Japanese Society for Information and Systems in Education, Vol.18, No.3, pp.297-307 (2001).

[9] M. Barak and S. Rafaeli, On-line question-posing and peer-assessment as means for web-based knowledge sharing in learning, International Journal of Human-Computer Studies, Vol.61, No.1, pp.84-103 (2004).

[10] Fu-Yun Yu, Yu-Hsin Liu and Tak-Wai Chuan, A web-based learning system for question posing and peer assessment, Innovations in Education and Teaching International, Vol.42, No.4, pp.337-348 (2005).

[11] N. Sugawara, K. Oda, H. Akaike, and H. Kakuda, SHoes: Realtime e-learning System for Classroom Education and Organizational Learning, Transactions of Information Processing Society of Japan, Vol.48, No.8, pp.2791-2801 (2007).

[12] Yuuki Hirai, Atsuo Hazeyama, A Learning Support System based on Question-posing and Its Evaluation, Proceedings of the Fifth International Conference on Creating, Connecting and Collaborating through Computing (C5), pp.180-185(2007).

[13] D. Hoshino, M. Takagi, N. Minami, and Y. Teshigawara, Navigation Function of Group Review for Promoting Collaborative Improvement of Quizzes Created by Students. The Eighth IASTED International Conference on Web-Based Education(WBE2009), pp.353-359 ( 2009)

# Effects of an Intuitional Pictograph Comment Function in a Video Sharing Web System

Kentaro Kagawa[*], Junko Itou [**], and Jun Munemori [**]

[*]Graduate School of Systems Engineering, Wakayama University, Japan
[**]Faculty of Systems Engineering, Wakayama University, Japan
{s105068, itou, munemori}@sys.wakayama-u.ac.jp

*Abstract* - Video sharing websites have spread throughout the world. Among the comments they contain are impression comments, which are one of the important factors determining the quality of video content. But if the posting process is complex or difficult, it is difficult to submit impression comments. Accordingly, we have developed a video sharing system named "Onion". One of the features of Onion is a pictogram comment function. The function consists of the scrolling wheel of a mouse and posting pictographs. We have experimented using the system and, as a result, obtained 13 videos, 738 views, 108 text comments, and 1,806 pictograph comments. The ratio of posted text comments is the same as before. The ratio of the posted pictograph comments is very large. We confirmed the utility of the system.

*Keywords*: video sharing, impression, comment, mouse wheel, pictograph

## 1 INTRODUCTION

In recent years, a great deal of video content has been shared owing to the enlargement of memory and hard disk drives, the spread of broadband, and the development of data compression technology [1]. There are many services using video content [2]. A video sharing website service is one such service. In particular, Nico Nico Douga [3] has become famous in the entertainment field. Users can post comments concerning particular video scenes. Comments include reviews, commentaries, impressions, dramatizations, and questions and answers. The actions of site users are divided into video search, viewing, and posting comments. But if it is complex and difficult for viewers to post comments, it becomes difficult to post their emotions or feelings to video. In this paper we propose an intuitional interface to post viewer's impression comments by using pictographs [4] and a mouse wheel device.

Chapter 2 explains the proposed video sharing system, and Chapter 3 shows the experiment. Chapter 4 describes the experiment results, and Chapter 5 is the conclusion.

## 2 PROPOSED MODEL

### 2.1 Composition of system

We have developed a video sharing website system called Onion. Table 1 shows a list of software that composes this proposed video sharing system. Figure 1 shows the software constitution of the system.

Table 1   Software constitution.

| component | software | version |
|---|---|---|
| Web server package | XAMPP | for Windows 1.6.6a |
| Web server | Apache | version 2.2 |
| RDBMS | MySQL | 5.0.51a |
| Scripting language | PHP | version 5.2.5 |
| View content | Flash | Professional 8 |
| Video encoder | FFmpeg | rev. 16905 |
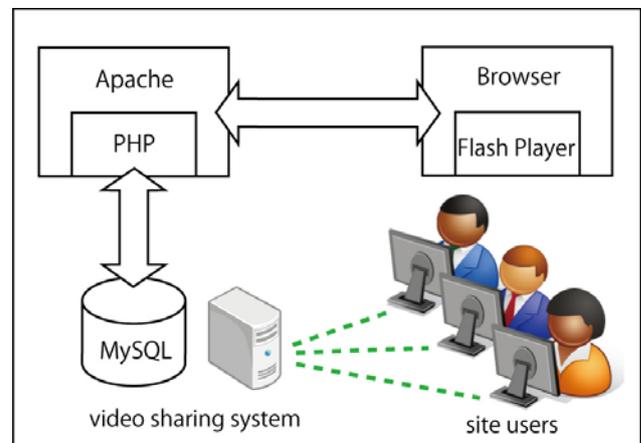


Figure 1: The constitution of the proposed system Onion

### 2.2 Function of system

The proposed system features an intuitional pictograph comment function. But the system also supplies some fundamental function services for site users like other video sharing systems. Figure 2 shows the top page of Onion.



Figure 2: Top page of Onion.

Site users can upload their original videos by using the video upload function with attendant information and authentication. If they wish to delete their videos, they can

delete them using the delete function with a password, which they had set when they uploaded it. Site users can also view uploaded videos by using the video search function and the video list. If a site user finds a video to view, the user then views the video on a viewing page. Figure 3 shows a content screen of a viewing page.


Figure 3: A content screen of a viewing page.

The video replay screen consists of a video screen, a text comment function, a pictograph comment function, and some additional functions. The content screen is 600px in width and height. The video replay screen is 600px in width and 400px in height (aspect ratio is 3:2).

Posted text and pictograph comments by video viewers are displayed on the video replay screen and flow from right to left. The size of text comments is 25px in height. The size of pictograph comments is 40px in width and height. The velocity of flowing is 150px/sec. Each comment is displayed on the video replay screen for 4 seconds. Viewers can post their comments by using the text comment function and the pictograph comment function.

## 2.3 Pictograph comment function

The process of posting a pictograph comment is done by selecting a pictograph and scrolling mouse wheel. The selected pictograph is chosen from the pictograph comment area that provides nine kinds of pictograph comments. Figure 4 shows the list of nine kinds of pictograph comments.


Figure 4: List of pictograph comments.

Each kind of pictograph can switch to three grades by scrolling the mouse wheel. Those grades depend on the number of revolutions of the mouse. Figure 5 shows the strength list of pictograph comments, and Table 2 shows the chosen pictograph strength by mouse scrolling.
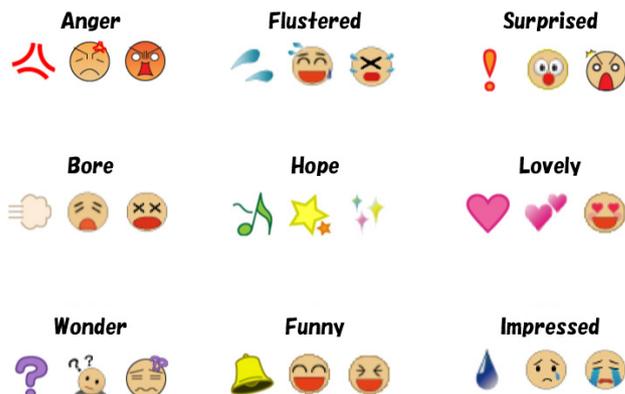

Figure 5: Strength list of pictograph comments.

Table 2: Chosen pictograph strength by mouse scrolling.

| impression strength | Amount of offset |
|---------------------|------------------|
| Strong | 25～ |
| Medium | 10～24 |
| Weak | 2～9 |

Each pictograph on the strength list is schematized starting from the left in increasing order. Some of the pictographs were created by Munemori's group [5],[6].

## 2.4 "Resonance Sense" function

Onion restricts pictographs to nine kinds. In other words, the frequency of posting the same pictograph comments is increasing. So, Onion provides a RS function (abbreviated form of the Resonance Sense). This function is designed to share emotions among viewers. The RS function is used when a viewer posts a pictograph comment. If the same pictograph comments are posted by others in the same video scene, the size of the just posted pictograph comment becomes large according to the number of like comments. The just posted comment expands 25px in width and height by each identical comment. The maximum pictograph size is 200px. At the same time, the other different comments fade out for a given length of time. Even though the same pictograph comments are posted more than two in the same video scene, RS function counts once at each posted. Figure 6 (a) shows a screen of not using the RS function (before), and Figure 6 (b) shows a screen of using the RS function (after).
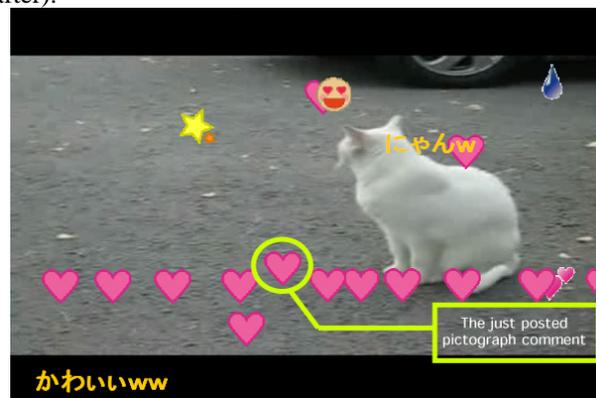

Figure 6 (a): A screen of not using the RS function (before).

Figure 6 (b): A screen of using the RS function (after).

# 3 EXPERIMENTS

We have carried out experiments by using the Onion system to prove utility of pictograph comments. The participants ranged from teenagers to those in their fifties in the experiments. Part of the experiments were carried out at a university festival. They were divided into 2 groups. One of the groups called the "view group" consisted of eighteen video view users (as viewers) and the other group called the "upload group" consisted of eleven video upload users (as uploaders). The view group included seven Wakayama university students and eleven members of the general public. The upload group consisted of Wakayama university students.

## 3.1 Material

For experiments, participants of the view group used computers equipped with a wheel mouse and were connected to 100 Mbps Ethernet LAN. The computers had sufficient speed to process streaming media and depicting screens.

The contents of video are shown below.
No.1: The automatic turn of a seat on a limited express
No.3: A vending machine with an interesting movement
No.10: An elevator
No.11: The backlashing of a picture
No.12: Playing in Shirahama
No.13: I'm breaking a watermelon
No.14: I'm eating sushi in large quantities
No.15: A dance show
No.16: Though I am lively, I am lonely
No.17: An encounter with a cat
No.18: Sculptures
No.19: Self-satisfaction
No.20: An analysis experiment of gum

## 3.2 Method

We shall now describe the experiment procedure of the view group and upload group participants. Figure 7 shows a scene of viewing a video (the view group).

### 3.2.1 Method for uploaders

(1) They uploaded some original videos beforehand using Onion.

(2) After the viewers posted comments, the uploaders checked the comments uploaded to their videos.

### 3.2.2 Method for viewers

(1) Search the randomly posted videos using Onion.
(2) View video on a content screen.
(3) Post comment using both comment functions.


Figure 7: A scene of viewing a video.

After the experiments, we distributed questionnaires to all the experiment participants.

# 4 RESULT AND DISCCUSION

## 4.1 Results of Experiments

The results of the experiments are shown below. We got 13 videos, 738 views, 108 text comments, and 1,806 pictograph comments. Figure 8 shows the number of the posted comments for each video.
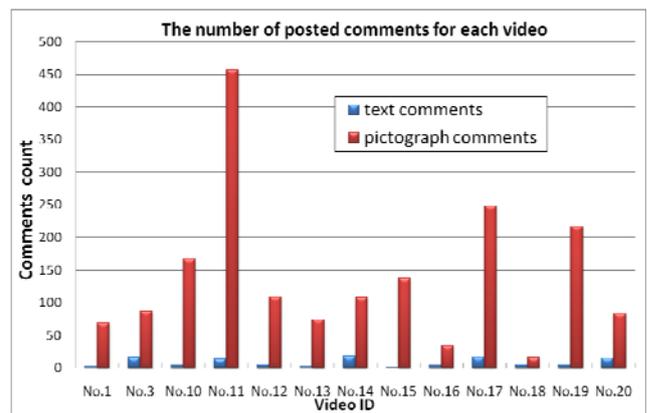

Figure 8: Number of posted comments.

The number of posted pictograph comments was more than the number of text comment for all uploaded videos. Posted pictograph comments accounted for 94% of the total posted comments, indicating that viewers prefer the pictograph comment function to the text one for expressing their emotions. Figure 9 shows the ratio of total pictograph comments. Figure 10 shows the number of posted comments for the three grades. Table 3 shows the number of posted comments in a viewing.
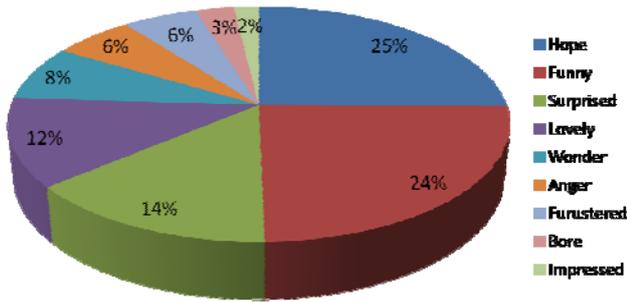
Figure 9: The ratio of total pictograph comments.

The ratio of posted pictograph comments expressing "Hopeful" was 25% and "Funny" was 24%. It means that these two kinds of pictograph comments occupied about half of the total pictograph comments. Conversely, the ratio of "Impressed" and "Bored" pictograph comments occupied only a small percentage. It indicates that some of the extreme or negative comments were posted less.
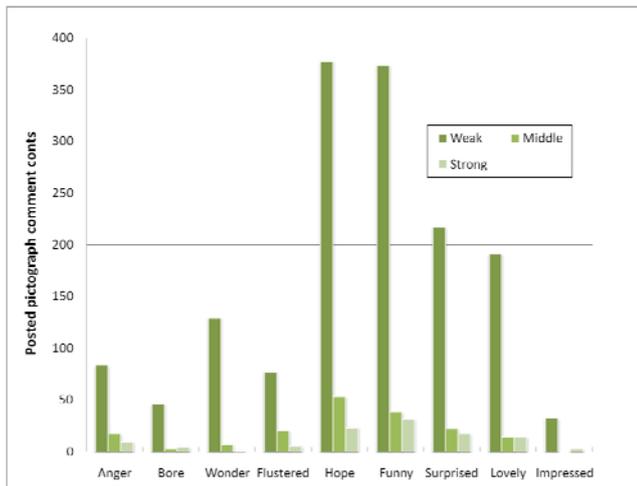


Figure 10: Pictograph comment counts for the three grades.

We got 1,527 weak grade pictographs, 173 middle grade pictographs and 106 strong grade pictographs. Thus, the weak grade pictograph comments occupied 85% of the total posted, indicating that the viewers prefer posting soft expressions to extreme expressions.

Table 3: The number of both posted comments

|  | Text Nico Nico Douga | Text Onion | Pictograph Onion |
|---|---|---|---|
| Average | 0.131 | 0.19 | 3.17 |
| Median | 0.134 | 0.123 | 2.631 |
| Expectation | 0.125 | 0.146 | 2.447 |

The number of total posted pictograph comments was 16.7 times that of text comments. There was no significant difference between the number of posted text comments of Onion and the number of posted text comments of Nico

Nico Douga. This indicates an increase of posted emotional comment counts.

Figure 11 shows the number of RS functions.
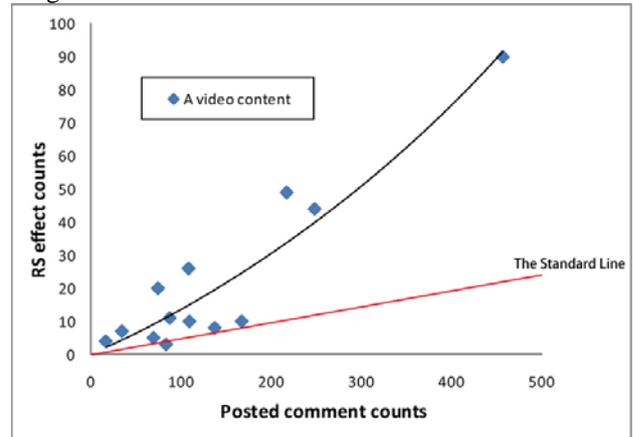


Figure 11: The number of RS functions.

The standard line indicates a value if each pictograph comments is posted evenly in all scenes. The standard line was calculated by values of average length of 11 videos (118.2sec).

The number of RS function occurrences was related to the number of posted comments. Therefore, the same pictograph comments that were posted by several viewers were concentrated in the same video scene. Then, the curve of the graph is above the standard line. It indicates that previously posted pictograph comments influence other viewers who watch the same video.

If they use the RS function, they may be able to get a sense of synchronization in disparate places.

## 4.2 Questionnaire results

### 4.2.1 Questionnaires for the upload group

Evaluations of part of the questionnaire were rated on a scale of one to five. "5" is the highest score and "1" is the lowest. Table 5 shows a part of the questionnaires for the upload group. The evaluation scores of the questionnaires were the average and the standard deviation of eleven video upload users.

Table 5: Questionnaire results for the upload group.

| Questionnaire items | Evaluation (AVG) / (STD) | |
|---|---|---|
|  | Text | Pictograph |
| Do you feel delightful if your videos get some comments from viewers? | (4.4)/(0.6) | (4.2)/(0.6) |
| Do you feel sad if your videos get no comments from viewers? | (4.1)/(1.0) | (3.8)/(1.1) |
| Do you have more incentive to next videos if your videos get comments? | (4.1)/(0.7) | (4.1)/(0.8) |

We carried out a T-test and found that there was little difference in evaluations between text and pictograph comments (Table 5), indicating no difference between the two functions. Thus, we can conclude that the pictograph comment function gave satisfaction for video upload users just like the text comment function.

### 4.2.2 Questionnaires for the view group

The results of the questionnaires for the view group may be summarized as follows.

1) Viewers can post whenever they wish.
2) Viewers do not have to "read" pictographs.
3) Viewers can grasp the comments of viewers at once.
4) Pictographs can be a distraction from watching video.
5) Range of expression is reduced.

The overall results indicate that pictograph comments have a high level of visibility and some problems (4, 5). So, pictograph comments require some method for solving these problems.

# 5 CONCLUSION

In this paper, we proposed an intuitional pictograph comment function for posting the emotions of viewers. The features of this function are using pictographs and a mouse wheel. This function was adopted in "Onion," a video sharing system we have developed. The pictographs of the system can switch through three grades by mouse scrolling.

We carried out experiments using "Onion" for twenty-nine participants who were divided into view and upload groups. We obtained 13 videos, 738 views, 108 text comments and 1,806 pictograph comments in the experiments. The results of the experiments indicated the following.

1) Viewers prefer the pictograph comment function to text one for posting their emotions.
2) Viewers prefer posting soft expressions to extreme expressions.
3) The same pictograph comments posted by several viewers concentrated in the same video scenes. The number of RS function occurrences were related to the number of posted comments.

Therefore, the proposed pictograph comment function is better than the text comment function for expressing spontaneous reactions.

In the future, we would like to discuss the video tags and java script of HTML, which are related our system.

# REFEENCES

[1] YouTube,Inc : YouTube, http://jp.youtube.com/

[2] D.Yamamoto, T. Masuda, S.Ohira, and K.Nagao, "Synvie: An Annotation System Based on Quotation of Video Scenes", Intraction 2007, pp. 11-18 (2007) (in Japanese).

[3] Nico Nico Douga, http://www.nicovideo.jp/

[4] Y. Ota, Pictogram Design, Kashiwa Shobou, Tokyo (1993) (in Japanese).

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

[6] 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)

# An Experimental Analysis of Accumulated Audiences' Comments for Video Summarization

Yoshia Saito [*], Yoshiki Isogai[*] and Yuko Murayama[*]

[*]Graduate School of Software and Information Science, Iwate Prefectural University, Japan
{y-saito, murayama}@iwate-pu.ac.jp, y.isogai@comm.soft.iwate-pu.ac.jp

*Abstract* - In this paper, we propose an audience-oriented video summarization scheme on video sharing services. The proposed scheme analyzes audiences' feedbacks such as rating and comments in a video and finds important scenes where there are a lot of feedbacks from the audiences. Then, the video is summarized by collecting the important scenes from audiences' point of view although typically it is summarized from video producers'/providers' point of view. As the first step toward the audience-oriented video summarization, we focus on comments as the audiences' feedbacks because currently some video sharing services allow audiences to comment on a specific scene storing their playback time. We assume there is a relationship between the number of audiences' comments on a scene and importance of the scene because the comments represent audiences' willingness to watch the scene. We report an experimental analysis for verification of the hypothesis and discuss some solutions to realize audience-oriented video summarization taking into account the experiment results.

*Keywords*: Internet broadcast, video sharing service, audiences' feedback, comments, audience-oriented video summarization.

## 1 INTRODUCTION

In recent years, most Internet users have broadband Internet connections and multimedia contents become popular on the Web. There are a lot of video sharing services nowadays such as YouTube [1] and Yahoo! Video [2]. A huge number of videos are shared and hundreds of thousands of new videos are uploaded every day. It is, however, difficult for audiences to find interesting videos quickly even if they retrieved dozens of candidates by appropriate keywords since it is required to watch the videos taking long time. A solution to the issue is to provide summarized videos.

Automatic generation of video summarization techniques have been studied by a lot of researchers [3-5]. In these studies, summarization is typically realized by understanding object and event in the video and selecting important scenes. Since these studies do not get directly feedbacks from audiences and there are a lot of audiences who have different feelings, it is difficult to keep interest factors of original video for the audience. To provide attractive summarized videos for the audiences, the video summarization should be audience-oriented. That means audiences' feedbacks should be applied to the video summarization algorithm to find scenes where the audiences get interested.

Meanwhile, most video sharing services have functions to receive feedback from audiences such as rating and comments. The received feedbacks are stored in a database and available for analysis of the videos. It would be possible to find scenes where the audiences pay attention by utilizing the feedbacks. Some video sharing services allow audiences to comment on a specific scene storing their playback time. Since each feedback is related with a specific scene, the feedbacks can be used as metadata about the scenes. Thus, current video sharing services already have good database to realize audience-oriented video summarization.

In this paper, we propose an audience-oriented video summarization scheme on video sharing services. The proposed scheme analyzes audiences' feedbacks in the video and finds scenes where there are a lot of feedbacks from the audiences. Then, the video is summarized by collecting the important scenes from audiences' point of view. As the first step toward the audience-oriented video summarization, we focus on audiences' comments as the audiences' feedbacks. We assume there is relationship between the number of audiences' comments on a scene and importance of the scene for video summarization because the comments represent audiences' willingness to watch the scene. To verify the assumption, we conduct an experiment collecting ten thousand comments per video from a video sharing service and discuss whether it is possible to make a summarized video utilizing the audiences' comments.

The remainder of this paper is organized as follows. In Section 2, we describe related work. Section 3 illustrates a model of audience-oriented video summarization on a video sharing service and describes a hypothesis. In Section 4, we conduct experiments for preliminary analysis and show the results. In Section 5, we discuss solutions to realize audience-oriented video summarization taking into account the experiment results. Section 6 gives some conclusions with a brief summary and future work.

## 2 RELATED WORK

We can save our time by summarized video and highlight video. Recently, we can also give feedback to watched videos and share our experience. In this section, we explain difference between the summarized video and highlight video and also describe scene extraction techniques which use audiences' feedbacks.

## 2.1 Summarization and Highlight

There is difference between summarization and highlight. We define the summarized video and highlight video by reference to typical researches [6-9] as follows:

- **Summarized video** shows the story of a video content in short time.

- **Highlight video** shows a set of interesting scenes of a video content in short time.

The motivation of our research is to provide short videos so that audiences can find objective video and grasp course of story of the videos quickly. We focus on the video summarization.

## 2.2 Audiences' Feedbacks

There are several scene extraction techniques which use audiences' feedbacks. In [10], audiences' browsing log such as "PLAY", "STOP", "PAUSE" and "JUMP" are used for the video summarization. The audiences unintentionally give their understanding of the video to the system through the browsing operations. They measure the subjective interestingness and importance using the browsing log. In sport videos, there is a technique [11] to use audiences' reactions such as cheering and applause. The proposed technique recognize audio signal in the sport videos and extracts interesting events for the video summarization.

A concept of time-tagging is proposed in [12]. Audiences can add time-tags to videos and these tags can be used as bookmarks. It is also applied to video summarization technique by analyzing the shared time-tags and scoring the tagged segments. Current video sharing and live streaming services provide feedback functions for audiences. Several video sharing services such as YouTube and Yahoo! Video has comment and rating functions. Audiences can submit text messages to the videos and rate the videos by 5-point scale. Most live video streaming services such as Ustream.tv [13] and Stickam [14] have a chat function. In these services, audiences can send chat messages among the audiences and its broadcaster in real-time. Nico Nico Douga [15] is a video sharing service in Japan and allows audiences to comment on a specific scene storing their playback time. The comments are displayed on the video field synchronized with the commented scene as if chatted with other audiences in real-time. Since the comments correspond with specific scenes and can be easily gotten them, we use the comment data in the Nico Nico Douga for our research.

## 3 AUDIENCE-ORIENTED VIDEO SUMMARIZATION SCHEME

The purpose of the audience-oriented video summarization is to provide summarized videos which keep interest factors of the original ones to audiences. In this paper, the "*audience-oriented*" means utilizing feedbacks from audiences as much as possible to provide a service from audiences' view of point. The audience-oriented service
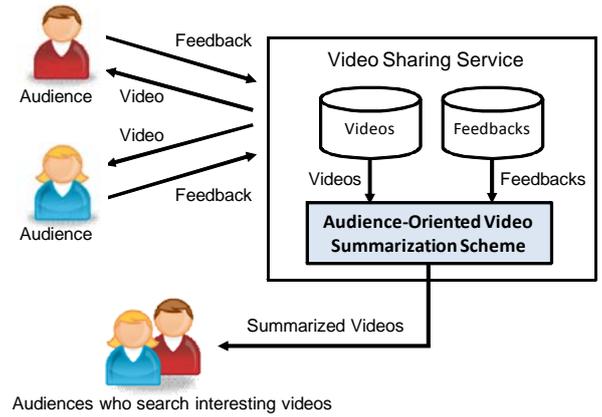


Figure 1: A model of video sharing service with audience-oriented video summarization scheme.

would improve audiences' satisfaction since it directly reflects the feedbacks.

## 3.1 Overview

Figure 1 shows a model of video sharing service with the audience-oriented video summarization scheme. In this service model, a service provider delivers videos to audiences and the audiences can give feedbacks to the service provider. The feedbacks are stored in a database of the service provider. When there are audiences who search interesting videos and have several candidates to watch, the service provider generate summarized videos of the candidates applying the feedbacks appropriately. The service provider offers the summarized videos to the audiences. The audiences can decide to watch a video by reference to the summarized video. If there are not enough audiences' feedbacks for the summarization, the videos are summarized by audio-visual video summarization techniques cooperatively.

## 3.2 Methodology

In this paper, we use audiences' comments which are associated to specific scene as the feedbacks. In order to study an algorithm for the audience-oriented video summarization, we have a simple hypothesis about relationship between video summarization and the audiences' comments. The hypothesis is as follows:

> *There is a relationship between number of audiences' comments and important scenes for the audiences. A scene which has sufficient number of comments is appropriate as a part of the summarized video.*

We assume audiences' comments increases when it is an important scene because the comments would represent audiences' willingness to watch the scene. The scenes which have a lot of comments would be worth watching for the other audiences and would be also important part of the summarized video. If the hypothesis is correct, we can get a set of candidate scenes for video summarization and
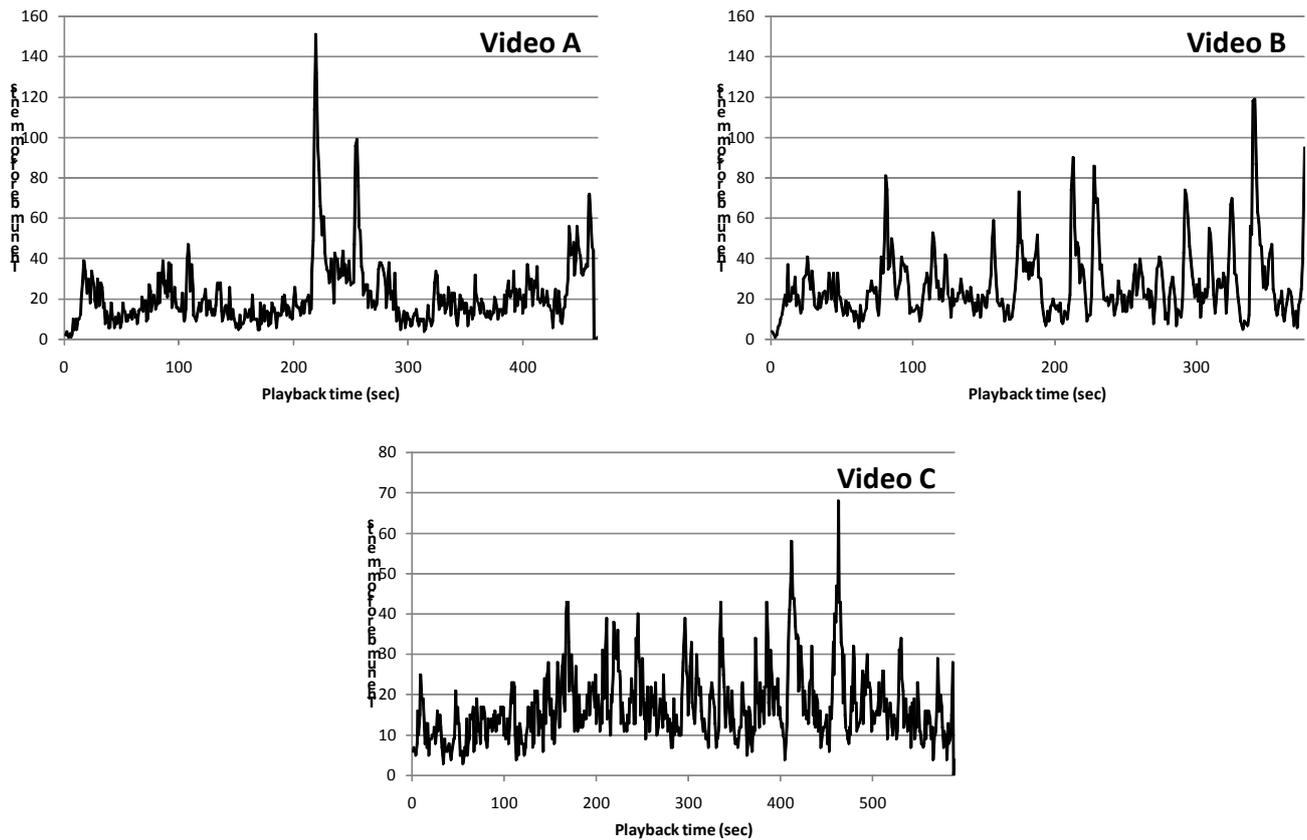
Figure 2: Changes in the number of comments per second.

generates the summarized video by putting several candidate scenes together.

Expected issues are that the highly-commented scenes are just interesting scenes for the audiences and they are not parts of the summarized scenes. In this case, the scenes are a set of candidates for highlight. We need to study the relationship between the number of comments and summarized/interesting scenes.

## 4  EXPERIMENTAL ANALYSIS

We conducted an experimental analysis to verify our hypothesis. For the analysis, we collected audiences' comments from a video sharing service and asked people to select scenes which are appreciate for summarized/interesting scenes. Then, we studied if the number of comments was positively correlated with summarized/interesting scenes.

### 4.1  Comment Collection

The comments data in the Nico Nico Douga is stored in a log database with the following information.

- Time and date when audiences commented
- Playback time when audiences commented
- User ID
- Comment

- Command to decorate the comment

We chose three popular videos (Video A, B and C) in Nico Nico Douga at random and collected ten thousand comments per video. The contents of the videos are as follows:

- **Video A**: A man makes a strange cake using a lot of cheep sweets and eats it. (Total length: 465 seconds)

- **Video B**: A man makes big balls of chocolate using a lot of small various chocolates and packages them. (Total length: 376 seconds)

- **Video C**: A man mixes various energy drinks and tries to drink the mixed one. (Total length: 589 seconds)

Each video has a story (introduction, making and completion). Figure 2 shows the changes in the number of comments per second. From the graph, high and low peaks can be clearly shown in each video. We presume these videos are suitable to verify our hypothesis and use them in the analysis.

### 4.2  Scene Selection

We asked 20 participants who are students in our university about the following questionnaire after watching each video.
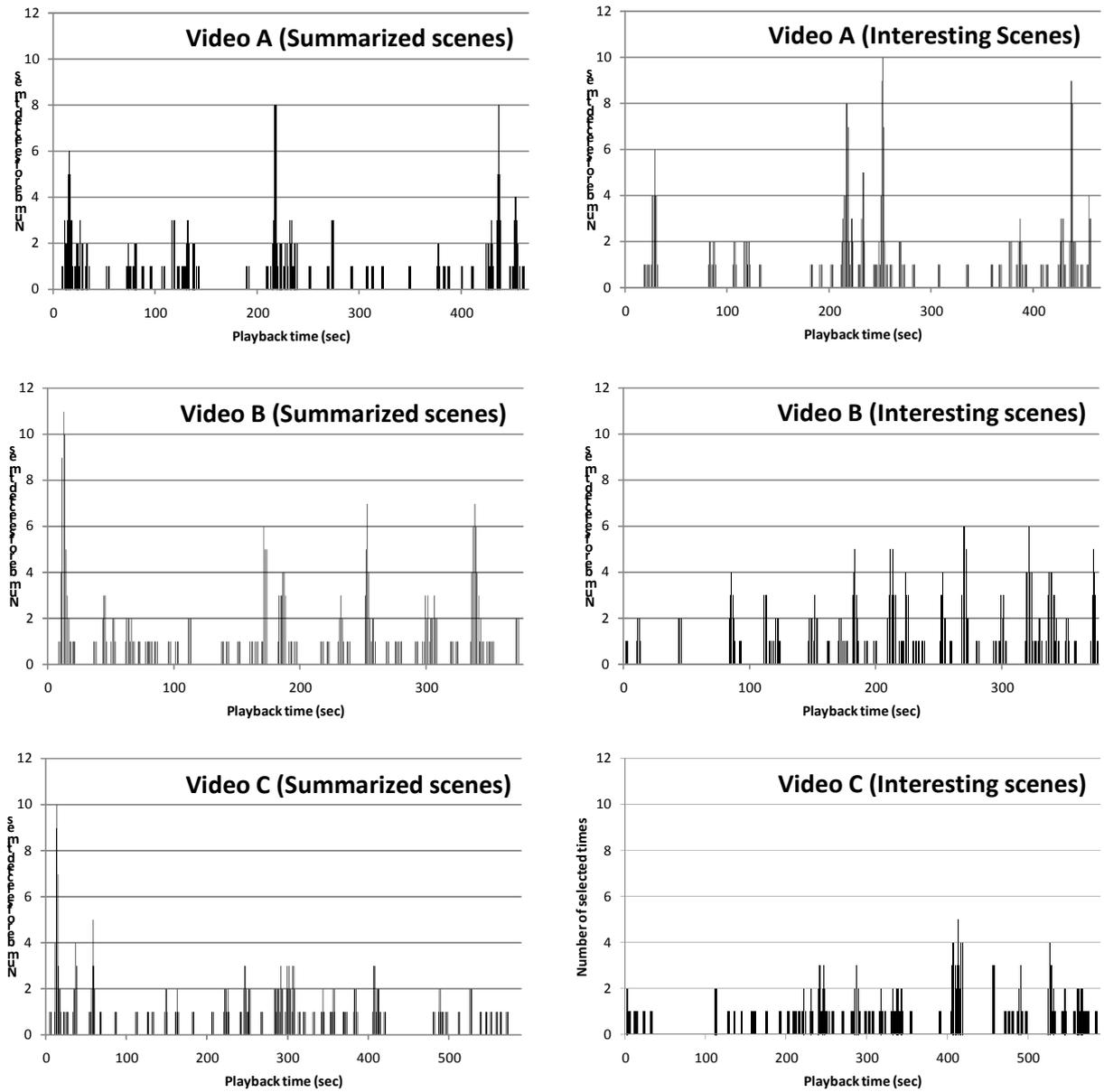
Figure 3: The number of selected times for summarized and interesting scenes.
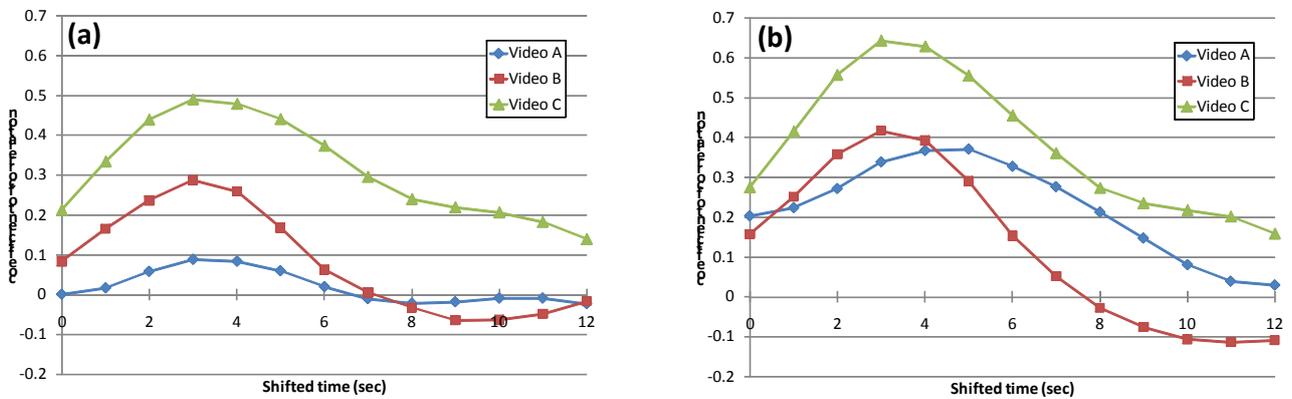


Figure 4: (a): coefficient of correlation between the number of comments and summarized scenes.
(b): coefficient of correlation between the number of comments and interesting scenes.

(Note: The order of watching the videos was at random for fairness)

1. Please select 5 scenes which are summarized the video on condition that each scene is 3 seconds.
2. Please select 5 scenes which are interesting in the video on condition that each scene is 3 seconds

After the questionnaire, we counted the selected times for summarized and interesting scenes. Figure 3 shows the results. We can see several differences between selected summarized scenes and interesting scenes in the results. In video A, there is an interesting scene around 400 seconds although it is not selected as a summarized scene. The scene shows interesting performance but it is not important to explain the story of the video. In the video B, we can see a scene which is selected as a summarized scene at the beginning of the video although it is not interesting. Because the scene shows the title of the content, it is selected despite it is not interesting. The same thing can be also said for the video C. At the beginning of the video C, a man explains the purpose of the video. Of course, it is not so interesting but important for summarization. Thus, we can see summarized scenes do not always correspond with interesting scene and a scene of introduction is important for video summarization even if it is not interesting one.

## 4.3 Analysis

We assessed coefficient of correlation between the number of comments and the number of selected times for summarized/interesting scenes. The result, however, does not show correlation between them. We presume that audiences would need to type their keyboard for a few seconds to comment to a scene and the input time should be required. Therefore, we shift the commented time to a few seconds before and assessed the coefficient of correlation again.

Figure 4 shows the result when the commented time is shifted by 1 second. From the graph, we can see the number of comments was positively correlated with summarized/interesting scenes when commented time was shifted to from 3 to 5 seconds before in these 3 videos. In the video A, the coefficient of correlation between the number of comments and summarized scenes is 0.09 when shifted to 3 seconds and 0.37 when shifted to 5 seconds as for interesting scenes. Weak correlation is shown only between the number of comments and interesting scenes. In the video B, the coefficient of correlation between the number of comments and summarized scenes is 0.28 when shifted to 3 seconds and 0.42 when shifted to 3 seconds as for interesting scenes. Weak correlation is shown between the number of comments and summarized scenes, and medium correlation as for interesting scenes. In the video C, the coefficient of correlation between the comments and summarized scenes is 0.49 when shifted to 3 seconds and 0.64 when shifted to 3 seconds as for interesting scenes. Medium correlation is shown between the number of comments and summarized/interesting scenes.

We found the number of comments was positively correlated with summarized/interesting scenes when the commented time was appropriately modulated in consideration of input time. The correlation strength differs in the contents of videos and the coefficient of correlation of interesting scenes is higher than that of summarized scenes.

## 5 DISCUSSION

The experimental analysis clarified summarized scenes do not always correspond with interesting scenes and the coefficient of correlation of interesting scenes is higher than that of summarized scenes. There are two issues. The first issue is how to extract a scene which is important for video summarization but few comments. The second issue is how to exclude scenes which have a lot of comments but inappropriate for summarized scenes. To solve the issues, we have two main approaches. The first approach is to make a support system which extracts candidate scenes using audiences' comments and suggests the scenes to users so that they can make a summarized video quickly and improve its quality. In this approach, the users decide whether the suggested scenes are appropriate or not and find missing scenes. The advantage of the first approach is ease of implementation and the drawback is workload of the users. The second approach is to devise an algorithm which finds unnecessary and missing candidates. We presume the number of comments is not sufficient as a parameter for the algorithm and additional parameters are required. For the additional parameter, meaning of the comments would be effective. Moreover, we probably need to use audio-visual summarization techniques together. The advantage of the second approach is to reduce human workloads and the drawback is difficulty of implementation. Since each approach has different advantages, we will study the two approaches as future work.

The experimental analysis also clarified commented time should be shifted to several seconds because of input time for comment messages. However, accurate time of the gap is not clear yet and we should estimate the gap time. One of the solutions is to focus on length of the comments and estimate the input time by multiplying average time for inputting one character by the length. The average input time would vary from person to person but it would be able to approximate the input time. In this case, we would have to take into account the combination of the inputted characters in order to estimate the input time more accurately.

Although we use collected ten thousand comments for the analysis in the experiment, the minimum number of comments required for extraction of summarized scenes should be discussed. Since there is no comment when a user uploads a video to a video sharing site, our proposed scheme cannot be applied and only audio-visual summarization techniques are effective. As time passes, audiences' comments are collected and our proposed scheme can be applied. By combining audience-driven summarization with audio-visual summarization, we presume the videos can be summarized more appropriately for audiences because it is difficult to know meaning of the scenes and audiences'

interests if there is only audio-visual information. We should study the threshold of number of comments to apply the audience-oriented video summarization by changing the number of comments.

## 6 CONCLUSION

In this paper, we proposed an audience-oriented video summarization scheme which analyzes audiences' feedbacks in the video and finds important scenes for video summarization in audiences' point of view. From the experimental analysis using audiences' comments in Nico Nico Douga, we got five findings; (1) summarized scenes do not always correspond with interesting scene, (2) a scene of introduction is important for video summarization even if it is not interesting one, (3) there is a short-time delay between comments and target scene, (4) the number of comments was positively correlated with summarized/interesting scenes when commented time was shifted to from 3 to 5 seconds before, (5) Some schemes would be required to make summarized video from audiences' comments because the audiences' comments indicated interesting scenes rather than summarized scenes.

As future work, we will design a support system for video summarization while studying an algorithm of video summarization based on the meaning of the comments so that we can generate summarized videos automatically. We will also compare the audience-driven video summarization method with some audio-visual summarization methods in order to show effectiveness of the proposed method more clearly.

## REFERENCES

[1]  YouTube, http://www.youtube.com/.

[2]  Yahoo! Video, http://video.yahoo.com/.

[3]  D. DeMenthon, V. Kobla and D. Doermann, "Video summarization by curve simplification", ACM Multimedia, pp. 211-218, 1998.

[4]  Y. Ma, L. Lu, H. Zhang, M. Li: A user attention model for video summarization, ACM Multimedia, pp. 533-542, 2002.

[5]  C. Kim and J. N. Hwang, "An integrated scheme for object-based video abstraction", ACM Multimedia, pp. 303-311, 2000.

[6]  B. T. Truong and S. Venkatesh, "Video abstraction: A systematic review and classification", TOMCCAP, Vol. 3, Issue 1, 2007.

[7]  H. Luo, Y. Gao, X. Xue, J. Peng and J. Fan, "Incorporating Feature Hierarchy and Boosting to Achieve More Effective Classifier Training and Concept-Oriented Video Summarization and Skimming", TOMCCAP, Vol. 4 Issue 1, 2008.

[8]  X. Tong, Q. Liu, Y. Zhang and H. Lu, "Highlight Ranking for Sports Video Browsing", ACM MULTIMEDIA, pp. 519-522, 2005.

[9]  M. Fleischman, B. Roy and D. Roy, "Temporal feature induction for baseball highlight classification", ACM MULTIMEDIA, pp. 333-336, 2007.

[10]  B. Yu, W.Y. Ma, K. Nahrstedt and H.J. Zhang, "Video summarization based on user log enhanced link analysis", ACM Multimedia, pp. 382-391, 2003.

[11]  Y. Rui, A. Gupta and A. Acero, "Automatically extracting highlights for TV Baseball programs", ACM Multimedia, pp. 105-115, 2000.

[12]  D. A. Shamma and R. Shaw, "Watch what I watch: using community activity to understand content", in Proc. of the international workshop on multimedia information retrieval, pp. 277-284, 2007.

[13]  Ustream.tv, http://www.ustream.tv/.

[14]  Stickam, http://www.stickam.com/.

[15]  Nico Nico Douga, http://www.nicovideo.jp/.

# Knowledge Resource Management

Kazuya Matsudaira*, Teruhisa Ichikawa**, and Tadanori Mizuno***

\* Graduate School of Science and Technology, Shizuoka University, Japan
\*\* Faculty of Informatics, Shizuoka University, Japan
\*\*\* Graduate School of Science and Technology, Shizuoka University, Japan
matsudaira@mizulab.net

*Abstract* - The last 50 years of computing can best be as "tumultuous". A variety of computers were introduced to solve different problems. As their capabilities grew, so did the types of applications. Each decade represented different milestone. The 1960's: The MIS movement; Mammoth information systems resolving around centralized mainframes. The 1970's: DataBase Management Systems; Designed to help integrate systems. The 1980's: Personal/Distributed Computing; End users could tap into the data-bases for their own purposes. The 1990's: Global Marketplace; Enterprises used information for strategic advantages in a global marketplace [1]. The 2000's: Click&Mortar; Cyber Enterprises dominate the way of business overcoming distances and national boundary. Although computers and types of applications have changed, the problems in application development have not. Those problems are not different than they were 50 years ago. If progress is to be made, governing rules and principles have to be established. A company must learn Knowledge Resource Management in order to move effectively into the year 2010 and beyond [2]. Knowledge Resource that is accumulated in a company consists of 4 components. The first one is Organization resource. The second one is Systems resource. The third one is Information resource. Data is the fourth one. Human being assigned to organization uses information through the system processed from raw material of data. Data is stored in either manual file or computerized file. Controlling these resources gives a company definitive strategic advantages〔3〕. *The paper proposes a framework of Knowledge Resource Management for enterprise's future wealth.*

*Keywords*: Knowledge Resource Management, Information Systems, IT, Data-Base, Organization Restructuring.

## 1  INTRODUCTION

Knowledge Resource Management represents an imaginative new way of thinking and conducting business in the Knowledge Age. It is not just another way of using computers and networks. Rather, it is concerned with innovation and management of all the resources required to create and utilize Knowledge. KRM will represent a radical departure from Information Age. The purpose of the paper is to convey to enterprises executives a practical approach for the resources consistent with their business objectives. Ultimately, the Knowledge Resource Management Approach demystifies their information systems process and put control back in the hands of executive management like Chief Knowledge Officer where it belongs. The concepts and ideas contained in the report created over thirty years of practices and experiences during the consulting jobs.
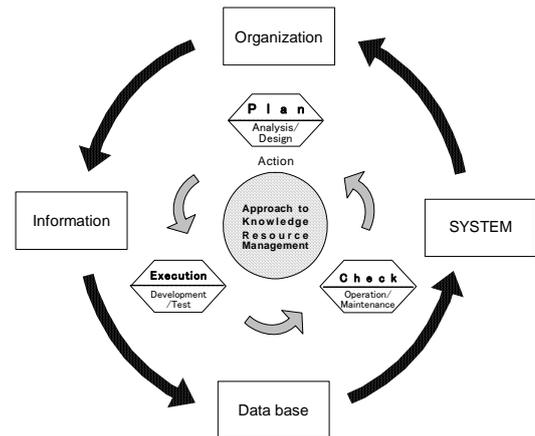


Fig.1 Approach to Knowledge Resource Management

## 2  BACKGROUND AND PROBLEM

Regrettably, in the past decade, attitude of computer orientation is still the "state of art" when developing system to solve business problems. In spite of the sophisticated technology available, the majority of enterprises still experience the following problems.

- Data-bases lack integration –The data elements used in the manufacturing system are not the same data elements used in the finance system. Each application jealously guards its data from shared use but other applications. As a consequence, data redundancy is rampant. A data element such as "part number" is typically defined and implemented differently for each application. This compounds problems when trying to implement changes to data on a consistent basis throughout an organization.

- Information requirements-Users at each organization level wants information for his/her decision-making on timing, but arrival of information is delayed and eventually useless. Users are estranged from IT department. Users want to talk about their business problems while IT experts want to impress everyone by talking technical jargons. The problems will become even more serious over time as end-users computing and personal computers proliferate.

- Systems are not documented-IT staff does not know where anything is. Systems users do not know how to use systems. Applications are difficult to maintain and modify. Consequently, systems require more man-hours than expected for keeping the systems up-

dated. In fact, due to poor documents of systems, it is often more expedient to completely rewrite the systems than to try to salvage existing investments with the current systems version.

· Organization are to be changed- Because of environmental changes, IT department is ready to respond for maintenance to adapt agile restructuring of organization. Adapting to the enterprise modifications, job descriptions, job manuals, rules and procedures must be modified.

· There is no control over the development process of organization, systems, information and data-base.

These problems are all to be traced back to the genuine belief.

# 3   PREDICTION

It is little surprising that there are not abundant research on the Knowledge of organization [4] [5]. Nonaka suggested SECI model for creation of Knowledge [6]. The day will come when management will recognize the importance of knowledge and learn how to apply it for strategic and tactical advantages. In order to use Knowledge resources in the same manner as other basic resources, such as financial, material and human resources, disciplined approach is required. In the environment of Enterprise Architecture, these Knowledge resources will be standardized for control and reuse among users of ERP packages. CKO is leading the change of enterprise, instead of CIO [7].

We predict a next generation after Knowledge Age will be Wisdom Age. Enterprises will experience significant prosperity. They will compete more aggressively and be able to change and move the organization at a speed unheard of in history. Their Information systems will be constructed based on the concept of Service Oriented Architecture. Evolutionally, a total new way of thinking and conducting business will be taken place (Table 1).

Table 1  The Four Stages of KRM Growth

| Stage & Name / Feature | 1. BIRTH | 2. CHILDHOOD | 3. ADOLESCENCE | 4. ADULTHOOD |
|---|---|---|---|---|
| | Data Age | Information Age | Knowledge Age | Wisdom Age |
| ENVIRONMENT | Data Processing | Decision Making | Enterprise Architecture | Service Oriented Architecture |
| APPLICATION | Bookkeeping & Calculating | Integrated MIS & DSS | ERP Based Systems | Service Based Systems |
| EQUIPMENT | Basic Computational Equipment | Centralized Computing | Decentralized Computing | Web Service Computing |
| PERSONNEL | ·Office Manager ·Clerical Staff ·Programmer | ·Chief Information Officer ·MIS Department ·System Engineer | ·Chief Knowledge Officer ·Project Office ·Knowledge Engineer | ·Chief Wisdom Officer ·Wisdom Center ·Service Scientist |

# 4   METHODOLOGY

Perhaps the clearest way to describe what Knowledge Resource Management is that Knowledge Resource Management is the management of the resources necessary to produce knowledge to satisfy business demands. The underlying rationale behind KRM is to inventory all of the resources required to develop Knowledge in the most efficient and cost-effective manner.

When standardized and controlled, these resources can be shared and reused by the employees and managements. This orientation to managing the parts of a product is very similar to the philosophy of Material Resource Planning (MRP) found in manufacturing field. In the manufacturing facility, we set the work stations and the mechanism of production control. Through front work stations, we design product, and following work stations are called assembly lines for assembling from parts to a product. Any product has a standard 4 levels of structure. Each level provides specifications for the next process. In this way, manufacturing process is standardized and preset all the necessary tasks from initial concept phase of a new product to final assembly phase. In same manner, we set up standards for each component as phases of resources development.

The first component of the Knowledge is organization. Organization has four level of structure. From top to the bottom, each is management level, director level, and middle manager level, and operator/clerk level.

System has same four level of structure. From top to the bottom, system level, sub-system level, administrative-procedure/computer procedure level and operation for human being and computer program for computer level.

Information has same four level of structure. From top to the bottom, Strategic information, Tactic information, Administrative information and Operational information are structured. Data has same four level of structure. From top to the bottom, Data-base, File, Record, and data-element are structured. The each component has engineered 10 phases to design and build and maintain it as resource.

Explosion/Implosion concept is engineering concept for resources utilization (Fig. 2).
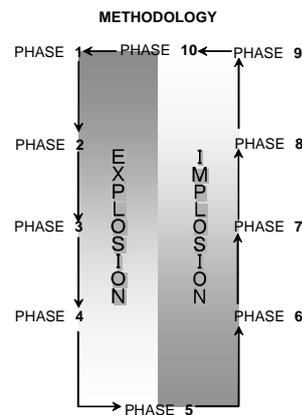


Fig. 2 Explosion/Implosion

Systems development methodology: PRIDE is well known. PRIDE methodology has 6 concepts and 10 phases [http://www.pmainstreet.com/mba/]. We suggest this methodology for system development and maintenance. Organization development methodology is created and proposed to apply it to become expert of Knowledge use. The methodology is consisting of 5 concepts and 10 phases. Both Information development methodology and Data-Base methodology which consist of 4 concepts and10 phases are recommended for use.

# 5   QUADRANGULAR PYRAMID MODEL FOR 4 COMPONENTS

Fig. 3 shows the model with internal relations between 4 components of knowledge resource. Model is for the concrete understanding Knowledge resources.

Any impact to the Organization component should be examined and analyzed through Systems component. After it, change impacts should be analyzed to Information component and Data-base component. Since those components are internally connected, from Systems components to Information and Data components change impact analysis should be done perfectly [8].
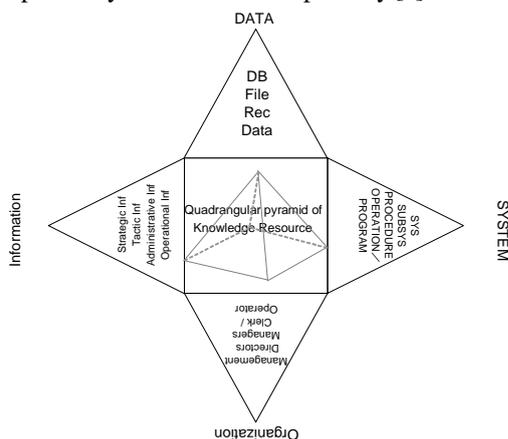


Fig. 3 Quadrangular pyramid of Knowledge Resource.

Whenever the enterprises see the sign of necessity for change, the Information requirements should be collected through out entire enterprise. Those definition sheets are to be analyzed and top management determines whether organization changes are needed or not. Then s/he decides type of organization pattern. Next to pattern selection among the types, s/he figures out concrete organization view pattern.

After these tasks, top management decides structures of enterprise using the collected sheets of Information Requirements. The last task is the drawing task of the organization charts. Without doing these tasks, Systems and Information and Data-Base should not be developed. Design or Redesign of Organization is the first task to do [9] [10].

# 6   INTEGRATED  USE  OF  4 METHODOLOGIES

Methodology must be applied to rationalize enterprise. To make it successful, integrated use of 4 methodologies are recommended. The way of use requires harmonized methods. Standard approach is that we should apply Organization

development methodology for Organization in the first. After it, Systems development methodology is to be applied. Next step is to use methodology for Information and/or Data-base [11].

Obviously, KRM is not a quick and dirty approach for solving managerial problems. Actually, it represents a cultural change for the entire enterprise, not just one area. Any enterprise has a culture: a way in which they behave and operate. Employees may be organized and disciplined by objectives which CEO has set up and elected to adopt. What is important is that in order for all employees to function and succeed, they must be able to recognize, accept and adapt to the culture. The decision on implementing the KRM methodology is a management issue, not just technical issue. Employees must first be convinced of its need.

Education on KRM concepts is necessary. Practices on Development of components like organization, systems,, information and database are inevitable.

A dictator approach may be fine if the dictator remain in power for some years.

Usually, the approach should be done with consensus by many of employees.

# 7   KNOWLEDGE   RESOURCE   MANAGER (KRM): THE TOOL FOR CONTROL

The Knowledge Resource Manager is the cohesive bond behind the Knowledge Resource Management concept. It represents the "bill of material processor" as illustrated in Fig. 4. It is the central storage where all of knowledge are inventoried and controlled. KRM software was developed by PJI in Japan [http://www.pji.co.jp].

To maintain the integrity of these resources definitions, they are cataloged in the KRM. This promotes sharing and reuse of all the resources by enterprise.
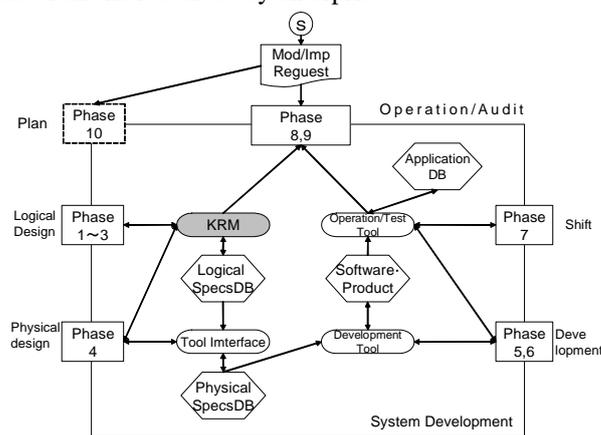


Fig. 4  KRM  availability

# 8   SUMMARY AND FUTURE WORKS

Since Knowledge Resource Management is new to all of enterprise CEOs, the concept might be misunderstood. It is not a computer jargon. It is an effective management methodology. CEOs must learn from the bench-mark company and try to apply it to make enterprise profitable. Eventually, it will impact strongly to the organization and

change will occur. Until the wealth of benefits is realized, CEOs had better keep patience [12]. Following are the some benefits gained by implementation of the methodologies.

- It is a practical solution based on common sense. It does not make use of cryptic terms or esoteric theories of design and management. The terminology and concepts are explicitly defined and related in plain Japanese. By doing so, it promotes communications and cooperation between users, developers and managements.

- It is a comprehensive methodology addressing all facets of Knowledge Resource Management, not just idea of researchers. The integration between methodologies represents a cohesive approach spanning from development functions to maintenance functions, thus providing continuity and cooperation.

- It provides a defined working environment that is fully documented, which provides means to expedite employees training. This means faster start–up of new employees, including clerks/operators, managers.

- It can be used to control and manipulate all the Knowledge resources. Consequently, resources can be quickly found and reused as required. Such reusability promotes a sharing Data-bases and integration of Systems. Further, inventory control provides the ability to perform "impact analysis" of a proposed resource changes. It also provides complete and current documentations of all organizational knowledge including systems structure in meaningful way.

It provides the means of effectively and efficiently utilization of knowledge resources, then human, money and material resources well utilized in an enterprise. It places the enterprise in the superior position.

To make our study more practically useful, we must prepare real examples as a reference and work sheets for smooth works. Also, we should develop education and training class for all levels of organization workers. It means we are going to complete our methodologies into the packaged product. We have already 4 actual implementation sites in Japan, those enterprises will bring more braving insights in a few years.

## REFERENCES

[1] H. Kitaoka, Business Intelligence, Toyo Keizai Inc., (2009).

[2] K. Sawka, You think you are so smart, Competitive Intelligence Magazine, Vol.1.6, No.2, pp.53-54 (2003).

[3] C. Argyris, Knowledge for action: A Guide to Overcoming Barrier to Organization Change, San Francisco, Jossy-Bass, (1993).

[4] J. Brown, P. Duguid, Organizing Knowledge, California Management Review, 40-3, Springer, pp90-111, (1998).

[5] N. M. Dixon, Common Knowledge, Harvard Business School Press in Boston,MA, (2000).

[6] I. Nonaka and N. Konno, Methodology of Knowledge Creation, Toyokeizaishipousya, (2007).

[7] I. Nonaka, Toward Middle-Up-Down Management: Accelerating Information Creation, Sloan Management Review, Vol.29-3, (1988).

[8] J. Botkin, Smart Business-How Knowledge Communities Can Revolutionize Your Company,The Free Press,a division of Simon & Schuster Inc., (1999).

[9] H. Mintzberg, Calculated Chaos, Harberd Business Review, Harvard Business School Press, (2003).

[10] L. Edvinsson, M.S. Malone, Intellectual Capital, Harper Collins Publishers, Inc., (1997).

[11] G. Hamel, Leading the Revolution, Harvard Business School Press in Boston, (2000).

[12] D.A. Nadler, Champion of Change-How CEOs and Their Companies are mastering the Skills of Radical Change, Jossey-Bass Inc., California, ( 1998).

# A proposal for a low-cost model of software process evaluation

Nobuhiro Kataoka *,     Tsukasa Kudou **

* School of Information and Telecommunication Eng., Tokai University, Japan
kataoka9@tokai.ac.jp
**Mitsubishi Electric Information Systems Corporation, Japan
kudou-tsukasa@mdis.co.jp

*Abstract* - Well-known process level evaluation methods used by software companies include CMMI certification and ISO/IEC TR 15504 process assessment, both of which are hearing methods with evidence base. These methods create a heavy workload for organizations that perform process evaluations as well as for the organizations being evaluated. However, since software companies collect various data in their everyday development work, we believe that the process level of such companies can be easily evaluated from these data, thus reducing the financial and other burdens of such evaluations. In this paper, we propose a low-cost model that uses these data to evaluate the process. The data used in the model include the specifications review man-hour rate, the test specifications review man-hour rate, and the software reuse rate. The model calculates the partial correlation coefficient in which the objective functions are productivity and number of bugs, and the explanatory variables are the three data items described above. Using this model, we can evaluate the amount of correlation from the results, and the strengths and weaknesses of an organization's process are revealed. We applied this model to an actual company's data and were able to determine the strengths and weaknesses of the company's software process from an evaluation of the results. Thus, the effectiveness of the proposed model in evaluating the software process at low cost was confirmed.

*Keywords*: software process, process evaluation, low cost, assessment, productivity

## 1   INTRODUCTION

Some well-known process level evaluation methods used by software companies include CMMI (Capability Maturity Model Integration)[1] certification and ISO/IEC TR 15504[2] process assessment. Both are hearing methods with evidence base. These create a heavy load for organizations that perform process evaluations as well as for the organizations being evaluated.

However, software companies in their everyday development work collect various data and we believe that the process level of software companies can be easily evaluated from these data, thus reducing the financial burden of such evaluations.
In this paper, we propose a low-cost model that uses these data to evaluate the process, specifically the effect of software reuse, design specifications review, and test specifications review, at low cost. In this way, we raise

awareness of these activities within the organization and also tie such awareness to making further improvements in the process.

Evaluation is performed by using the value of the partial correlation coefficient in which the objective functions are productivity and number of bugs present, and the explanatory variables are the specifications review man-hour rate, the test specifications review man-hour rate, and the software reuse rate. When there is an area that does not produce a good effect, this indicates where in the software development process a problem exists.

In this way, the weak points in the software development process can be determined using our model. As a next step, more detailed analysis is required to pinpoint the exact areas for process improvement.

To determine whether the proposed model was effective for an initial general evaluation of the process at low cost, we used actual system data to evaluate the model. As a result, we were able to determine the strengths and weaknesses of the organization's process, thereby confirming the effectiveness of our model.

The remainder of this paper is organized as follows. Section 2 presents the research situation and current challenges in process evaluation, and Section 3 introduces the proposed model. Section 4 describes the results when actual system data were applied to the proposed model, Section 5 presents the overall evaluation results, Section 6 discusses the significance of our model, and Section 7 provides concluding remarks.

## 2   METHOD OF SOFTWARE PROCESS EVALUATION

Simply acquiring ISO 9001[3] certification does not improve the software development process of an organization. Rather, process improvement is carried out on a daily basis using a quality system that has been developed over time. In addition, to acquire CMMI certification, the process level is certified according to the level of the certification acquired, but great expense and effort is necessary for this. The software development process can otherwise be evaluated by the process assessment of ISO/IEC TR 15504, however this also requires much effort.

It is well known that upstream review plays a considerable role in improving the quality of software [4].
A method has been proposed to improve the efficiency of review methods [5], and there is also a proposal on how to zero-in on a process that has become a problem [6]. In addition, Komuro et al. collected three data items-review

speed, indication density, and review efficiency-and, using a statistical technique, identified problems and made improvements, confirming the improvement of the review effect and cost reduction [7].

All these improvement methods are for organizations that have processes of a certain level, and only focus on reviews that have a strong relationship to quality. However, there are many organizations worldwide that do not reach this level. In such organizations, advancing to the next step comes only after analyzing the strengths and weaknesses of the existing processes of the organization. Thus, a low-cost process diagnosis model requiring only collected data is necessary.

# 3 PROPOSED MODEL

We used the following three data items in our proposed model: 1) design specifications review man-hour rate; 2) test specifications review man-hour rate; and 3) source code reuse rate. The data we used as evaluation result data were the development productivity and the bug rate (number of bugs per line of source code). Our reasons for choosing these data were as follows. The importance of review in software development is well known. Software development is human intellectual work, and there are many elements in which we may make errors, for example, through a lack of technical skills, incorrect beliefs, or discrepancy with the work of other people.

In such a situation, 7.3.1 b for ISO 9001-2000 requires that a review suited for each stage be performed, and the inspection propriety confirmed. In addition, CMMI requires a review known as "a pure review" in the third level key process area.

On the other hand, the reuse of source code has been said to be the trump card of productivity improvement. It is natural that productivity will improve with code reuse, but the impact on productivity changes greatly depending on the type of code reuse. In other words, the impact on productivity and quality changes greatly depend on whether it is reuse of code at source level or whether it is reuse of modules. Recently, a number of tools that automatically generate source code have been developed. However, only the skeleton of the source code is generated by design specifications and the details must be added by the user. In addition, due to the popularization of Java, the use of modules is spreading with the use of JDK and Java Beans, but application logic for domains are limited to reuse among sections. Moreover, in certain domains the product is used as a framework after system development is complete, and reuse is planned by changing some part of the system, thus improving the productivity.

## 3.1 The development model

Figure 1 presents the development model which is the premise of the proposed model. This development model is a general waterfall model, but we will describe it in order to eliminate any possible misunderstanding. In the figure, we show the flow of program development, the stage where the program reuse process is performed, and the stage where the review is done.

In addition, we clarify the meaning of productivity of the overall software, productivity of program development, total bugs, combinatorial testing bugs, and system testing bugs.

## 3.2 An evaluation model

(1) In the evaluation model, we assume that the following data are collected at each project.
 - All of the software development lines of code
 - New production lines of code
 - Reused lines of code
 - Man-hours throughout the entire development
 - Man-hours creating program specifications + man-hours creating the program + man-hours testing modules (hereafter, program development man-hours)
 - Source code reuse rate
 - Man-hours used for the specifications review
 - Man-hours used for the test specifications review
 - Number of bugs in combinatorial testing
 - Number of bugs in system testing (including field testing)
 - Man-hours for bug repair from combinatorial testing to field testing (bug repair man-hours)

We selected these data items assuming that many software companies collect them. Of importance in our model is that it can perform analysis even if some data are absent without impacting the overall situation, and it can also watch over wider overall tendencies. The reason we did not include review man-hours of the source code in the model is that we determined that the numbers of cases in which the source was reviewed were few, with the exception of specific parts such as important parts of the system or parts made by new employees.

(2) Data classification and collection
Here, we describe how to create the evaluation data from the above-mentioned data
 - Total productivity: Lines of source code from the entire software development / man-hours used throughout the entire development
 - New total productivity: Lines of source code from the new production / man-hours used throughout the entire development
 - Program development productivity: Lines of source code from the entire software development/ man-hours used throughout the program development
 - New program productivity: Lines of source code from the new development/ man-hours used in new partial program development
 - Specifications review man-hour rate: (system specifications + software specifications + program specifications) man-hours used in review / man-hours used throughout the entire development
 - Total man-hours used for bug repair: Bugs repair man-hours used in combination testing, system testing and field testing
 - Number of bugs in combinational testing: Number of bugs in combinational testing
 - Number of bugs in system testing: Number of bugs in system and field testing.

a. whole software productivity          b. program development productivity
c. total bugs    d. combination testing bugs    e. system testing bugs
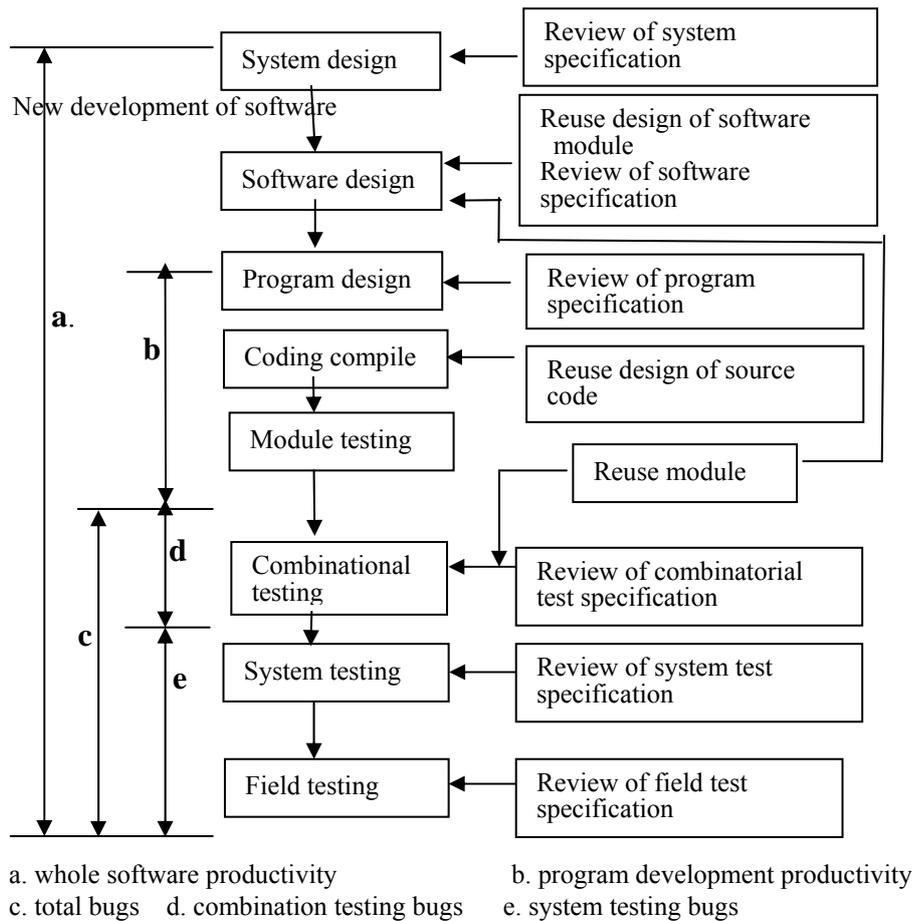
Fig.1    Process of software development

## 3.3 Desirable value for the model

The proposed model calculates the partial correlation coefficient in which the objective functions are productivity and number of bugs present, and the explanatory variables are specifications review man-hour rate, test specifications review man-hour rate, and software reuse rate. The number of bugs is normalized by dividing it by a development line of source code. It is necessary to use partial correlation analysis because in simple correlation other elements are impacted, and it is not possible to perform evaluation of the exact degree of impact. As a result, we believe that the desirable tendency is as shown in Table 1. The table is explained as follows.

(1)  Source code reuse rate:

It was natural that the total productivity and source code reuse rate would show a high correlation, so we set this value to 1.0. For the productivity of a new development part, there is a system framework, and it shows high correlation if we have a style change or addition for the framework. However, because all development is not done in this style, we cannot expect it to be 1.0. Therefore, in Table 1 we set this value slightly lower than 1.0, at 0.8 as a temporary value.

It is natural that entire program productivity show a high correlation, so we set this to 1.0 We set the correlation value to 0 for real new parts which will have no correlation.

Because bugs are not included if the reuse of source code is at the program level or module level, bug repair man-hours might be in high negative correlation. In addition, the number of bugs in combinatorial testing and system testing may have a high negative correlation for similar reasons. In the case of reuse of source level code, it is very likely that bugs will be created in change stages and insertion stages even if bugs were not in the original source code, so there will not always be a high negative correlation. Therefore, we temporarily set a low value of -0.8 rather than a value of -1.0 here.

(2) Specifications review man-hour rate:

When the review man-hour rate is high, an improvement in productivity and a reduction effect in the number of bugs are recognized for the following reasons.

Here, we are assuming an organization with a 5% or lower review man-hour rate, not organizations with more than a 10% man-hour rate.

**Table 1　Desired value of this model**

| | Total development productivity | | Program productivity | | Bug repair total man-hours | Number of bugs | |
|---|---|---|---|---|---|---|---|
| | Total | New part | Total | New part | | Combinational testing | System testing |
| Software reuse rate | 1.0 | 0.8 | 1.0 | 0 | −0.8 | −0.8 | −0.8 |
| Specifications review man-hour rate | 1.0 | 1.0 | 1.0 | 1.0 | −0.2* | −0.2* | −0.2* |
| Test specifications review man-hour rate | 0.8** | 0.8** | 0.8** | 0.8** | −1.0 | 1.0 | −1.0 |

* We expected some negative correlation and so used - 0.2.
** We expected considerable positive correlation and so used 0.8.

Total productivity, new part productivity, the entire program productivity, and new program productivity should show a high correlation due to the reduction of rework man-hours by the review man-hour rate of system specifications, software specifications, and program specifications.

Therefore, we set this value to 1.0. Likewise, program productivity should show a strong positive correlation with the review man-hours rate. This is similar for the entire program and new parts. Therefore, we set this value to 1.0.

Because errors are extracted by the specifications review, the bug repair man-hours will show a negative correlation. In the specifications review it is difficult to perform a review with enough detail that enable program errors to be extracted, so we cannot expect a large value as negative correlation. Therefore, this is temporarily set to -0.2. We temporarily set combinatorial testing and the number of bugs of system testing to - 0.2 for similar reasons.
(3) Test specifications review man-hour rate:
If the test specifications review man-hour rate is high, the bugs in combinatorial testing increase exponentially and will show a strong positive correlation with combinatorial testing. Therefore, we set this value to 1.0. On the other hand, by detecting a bug by combinatorial testing early, the number of bugs can be reduced in the system testing plus field testing (hereafter, "system testing"), so we expect a strong

negative correlation. Therefore, we set this value to -1.0. In addition, the repair costs for a bug are large in system testing for combinatorial testing, so a strong negative correlation was expected for the bug repair cost. Therefore, we set this correlative value to -1.0. Because the bug repair cost decreases, a strong correlation with total productivity can be expected.

However, this correlation is not, in actuality as strong as anticipated. In addition, it is similar to the productivity of total and new development parts. Therefore, we set this value to 0.8. Because we can say that the productivity of program development will be similar, also we set this value to 0.8.

## 3.4 Contribution rate for total productivity

A characteristic of this model is the fact that we can evaluate a single year of data. However, it can be used to evaluate the contributing causes for improvements in productivity.

For example, we assume that $\alpha$ is the productivity improvement rate for the current year over the previous year. "A" is the growth of the source code reuse rate, "B" is the growth of the design specifications review man-hour rate, and "C" is the growth of the test specifications review man-hour rate.

In addition, we set each correlation value for total productivity as a, b, and c, and set each contribution rate for each productivity as $\beta a$, $\beta b$, and $\beta c$. First we find $\beta a'$, $\beta b'$, and $\beta c'$, which are the apparent contribution rates for each element. This can be expressed by Equation 1:

$$\beta a' = A \times \frac{a}{a+b+c}$$
$$\beta b' = B \times \frac{b}{a+b+c} \qquad (1)$$
$$\beta c' = C \times \frac{c}{a+b+c}$$

Next, we distribute the total contribution rate using the apparent contribution rate, and calculate the actual contribution rate. This can be expressed by Equation 2:

$$\beta a = \frac{\alpha \times \beta a'}{\beta a' + \beta b' + \beta c'}$$
$$\beta b = \frac{\alpha \times \beta b'}{\beta a' + \beta b' + \beta c'} \qquad (2)$$
$$\beta c = \frac{\alpha \times \beta c'}{\beta a' + \beta b' + \beta c'}$$

The following shows a calculation result using a data sample.

$\alpha$ =20%  A=30%  B=10%  C=15%
a=0.8   b= 0.9   c=0.85
$\beta$ a'=30×0.8/2.55= 9.4 %
$\beta$ b'=10×0.9/2.55= 3.5 %
$\beta$ c'=15×0.85/2.55= 5.0 %

$\beta$ a =20%×9.4/(9.4+3.5+5.0)=10.5 %
$\beta$ b =20%×3.5/(9.4+3.5+5.0)=3.9 %
$\beta$ c =20%×5.0/(9.4+3.5+5.0)=5.6 %

## 4  EVALUATION OF MODEL USING DATA FROM AN ACTUAL COMPANY

We evaluated the model using data from an actual company's projects. The data was derived from 30 projects carried out during a certain year, and the development number of source code lines varied from 5KL to 700KL, with an average of 120KL. In addition, reuse rate in the domain concerned was considerably high at about 50% on average, and the highest projects were at 90%. Of course, there were also projects at 0%.

The results from an evaluation of these data using our model are shown in Table 2. Next, we will describe the analysis results. In addition, we show the desired value of the model and the real data in a radar chart in Figure 2. However, we reversed the positives and negatives in this chart because a decrease in the number of bugs is desirable. However, since an increase in the specifications review man-hour rate is desirable, we do not reverse the positives and negatives here.

(1) Correlation with source code reuse rate
The correlation between total productivity and source code reuse rate is close to the desired value. In addition, the correlation of the productivity of a new development part with the source code reuse rate is also close to the desired value. This means that when there is a framework for the entire system, repetitive development is done, such as adding new functionality. As a reuse method, we can determine that this is a desirable development process.

The program production correlation is lower than the impact of total productivity.

As for this, reuse of the source code for reuse of the design specifications might require extra work. We believe this is because it is reuse of the source code itself, and not reuse of parts of the source code. If parts of the source code are reused, the correlation should be of the same degree as the correlation with total productivity. There is no correlation with the productivity of new parts of program productivity. This is in accordance with the expected value.

Bug repair man-hours have a value close to the expected value. We believe this to be caused by the reduction in the number of system test bugs. However, in regards to the number of bugs in combinatorial testing and system testing, they were considerably different from the expected values. We believe that this is because the source code was reused, and it was not just reuse involving most of the system or parts of the system.

(2) Correlation with specifications review rate
In total development productivity, the total productivity and new partial productivity both had a correlative value close to the expected value. We believe that a sound design review was performed for the system specifications and software specifications. We believe that the correlation of new parts was particularly high because many items are pointed out for design review because they are new, so the review effect is high.

For program productivity, total productivity displays a considerably high correlation. For new parts, the correlation is considerably lower than the expected value. We believe that a sound review could not be done unless someone who understands the contents of the program specifications performs the review.

There is no clear correlation with number of bugs, and the results do not accord with the expected values at all. This means that the review of the program specifications was not done to the error extraction level. This suggests that, in order to extract a program bug, someone who understands the program must do the review of the program specifications, or a review of the source code must be done.

(3) Effect of the test specifications review
The review rate of specifications has some correlation with total productivity. We believe that it has the effect of reducing the number of bugs repair man-hours. However, no correlation is observed with new production parts.

Table 2: Example of estimate of partial correlation coefficient using actual company data

| | Total development productivity | | Program productivity | | Bug repair total man-hours | Number of bugs | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | Total | New part | Total | New part | | Combinational testing | System testing |
| Software reuse rate | 0.94* | 0.87* | 0.90* | 0.09 | −0.82* | −0.44 | −0.54 |
| Specifications review man-hour rate | 0.92* | 0.95* | 0.88* | 0.69 | 0.20 | 0.14 | 0.30 |
| Test specifications review man-hour rate | 0.71 | 0.32 | 0.73 | 0.53 | −0.81* | 0.84 | −0.35 |

*    shaded areas denote statistically significant results

**Resue rate**

Legend: Actual software reuse rate; Desired software reuse rate

Axes: Total productivity, Productivity of new part, Total productivity of program, Productivity of new program, Bugs repair total man-hours, Number of bugs combinational test, Number of bugs system test

Scale: 1.00, 0.90, 0.80, 0.70, 0.60, 0.50, 0.40, 0.30, 0.20, 0.10, 0.00

**Specifications review**

Legend: Actual specifications review man-hour rate; Desired specifications review man-hour ratio

Axes: Total productivity, Productivity of new part, Total productivity of program, Productivity of new program, Bugs repair total man-hours, Number of bugs combinational test, Number of bugs system test

Scale: 1.00, 0.80, 0.60, 0.40, 0.20, 0.00, 0.20, 0.40

**Test specifications review**

Legend: Actual test specifications review man-hour rate; Desired test specifications review man-hour rate

Axes: Total productivity, Productivity of new part, Total productivity of program, Productivity of new program, Bugs repair total man-hours, Number of bugs combinational test, Number of bugs system test

Scale: 1.00, 0.90, 0.80, 0.70, 0.60, 0.50, 0.40, 0.30, 0.20, 0.10, 0.00

Fig. 2  Radar chart of desired and actual partial correlation coefficient

It is easy to carry out a review of specifications for existing parts with a fixed architecture, whereas for new parts, it is difficult to do a review of test specifications. It is similar in the case of the correlation with program productivity.

In regards to the correlation with bugs, man-hours used to repair all bugs shows a considerable negative tendency and, in this aspect, is close to the expected value. In addition, the number in combinatorial testing shows a high positive correlation, and this is in accordance with the tendency of the expected value. However, the number of bugs in system testing shows a negative correlation tendency, but the value is insufficient. This shows that the test specifications review is not yet sufficient, and as a result, the bugs could not be completely caught in combinational testing.

(4) Contribution rate for total productivity
In this case there was no data on the productivity improvement rate from the previous year so were not able to perform an evaluation. However, in Section 3.4 we showed that a, b, and c for the reuse rate, specifications review man-hour rate, and test specifications review man-hour rate were 0.94, 0.92, and 0.71, respectively. Therefore, we presume that entire productivity is probably considerably dependent on the first two.

## 5. EVALUATION

From the analysis we could analyze the strengths, weaknesses, and areas for improvement for this organization, thereby confirming the effectiveness of our model. The results can be summarized as follows.

(1) Strengths in the organization's process
Reuse of software depends on the design specifications framework, and this is reused for system designs and software designs. By further promoting this in the future, we can expect higher productivity and quality improvements from software reuse. Specifications reviews are largely being done in the upstream process. Test specifications reviews have a considerable effect on reducing the number of bug repair man-hours

(2) Weaknesses in the organization's process
We believe the following are necessary process improvement items. In software reuse, the reuse must be approached not in terms of source code, but in terms of parts, or as a framework. Downstream process reviews such as program specifications reviews do not show a large enough effect. Specifically, they do not have an effect on bug extraction. We believe that there are a few review effects for new programs in particular and the following items are necessary to efficiently review a new program: a review checklist, setting the numerical quota of bug search items, and an efficient explanation of the specifications by the creator. A checklist is also necessary for an efficient specifications explanation, and performing a prior check based on this checklist is also one method.

In test specifications reviews, a more thorough review of newly produced parts is required. This also requires a review checklist, setting numerical quota for bug search items, and an efficient explanation of the specifications by the creator.

(3) Effectiveness of the model
From the previously mentioned analysis results, we were able to make an analysis of the overall strengths and weaknesses of the organization. In addition, as a result of a hearing undertaken with the development section, we heard the opinion from many that some of the items were vague. Presenting items with a clear number makes it easy for managers to choose the necessary action, and it increases their sensitivity towards developers. This shows the effectiveness of this mode

## 6 DISCUSSION

Detailed data was collected on the organization chosen for evaluation and the various data were consolidated for model evaluation. We believe that there are a number of organizations that collect these data but do not use it in this way, and our model is likely to be beneficial for such organizations.

As a next step, it is necessary for them to take measures to improve their weak points. While various methods have been proposed, it is necessary to adopt an improvement method according to the organization's specific situation [8][9][10][11]. In addition, it is necessary to extract problem processes from upstream specification review results. This will cause the effectiveness of the review to increase.

We have previously proposed such a method [6]. With this method, when upstream process management is insufficient, we focus our attention on the fact that the effect could appear in the each bug reports of the test process in the downstream processes. We can analyze the bug reports and identify groups that we believe to have characteristics similar to the process impact. Then, by following the process characteristics noted in ISO/IEC 15504, and the combinatorial table from the work output, we can determine the particular upstream processes that were the cause of most of the bugs occurring in the downstream tests and then make improvements.

Also, in this example, because the man-hours used in the upstream process review were comparatively low (less than 5%), the review man-hour rate and productivity had an extremely positive correlation. If the review man-hour rate becomes high, the challenge will be how to conduct reviews efficiently. We have also previously suggested a method to improve review efficiency [5] in which a quality coordinate [7] is initially used to set a start and end standard point of review. Following on form this, we have proposed here a method to continuously improve the review checklist using the statistics technique of a management diagram. As a result of applying this method to an actual organization, we were able to confirm the tolerance of the quality coordinate, the review cost, and a shift in the test man-hours to the upstream process for the entire lifecycle. In addition, the product quality is gradually improving. Improvements were confirmed for each criterion, implying the effectiveness of the proposed method.

# 7    CONCLUSIONS

The characteristics of this model are, firstly, it is able to make a process evaluation using the data collected every day within an organization and, secondly, even a single year of data can be used to evaluate the process, multiple years are not needed. Although this is only analyzing general tendencies, there is much value in using the model because analysis can be done using data that is collected daily, and thus it requires less labor and time than assessments done by outside companies or quality management organizations.

Using our model, we can find the weaknesses of organizations and improve them, and we can connect this to actions to make the strengths even stronger, clarifying the direction for such organizations. For future work, we will apply this method to other projects in order verify whether it has a similar effect.

## REFERENCES

[1] http://www.sei.cmu.edu/cmmi/index.html
    (2009/8/20)
[2] http://www.isospice.com/   (2009/8/20)
[3] ISO 9001:2000
http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=46486   (2009/8/20)
[4]   Rogger S.Pressman, "SOFTWARE ENGINEERING /sixth edition"  McGraw-Hill (2005)
[5] A. Hayashi, N. Kataoka, "A Method to Identify Critical Software Process Improvement Area using Quality Function Deployment". Proceeding International Conference on Innovation in Software Engineering   ISE, pp1143-1148 (2008)
[6]  A. Hayashi, N. Kataoka , "A Method to Identify Critical Software Process Improvement Area using Quality Function Deployment",       Journal C of   IEEJ , Vol.128, No7 , pp1231-1241 (2008)
[7] Komuro et al. "Peer review technique improvement and quantitative analysis of improvement effect on the basis of the reality of the development work " ,SEC Journal, Vol.1, No.4, PP.6-15, Nov. 4(2005)
[8] Tom Gilb, D. Graham, "Software Inspection", Addison-Wesley Professional (1993)
[9] Peter C. Rigby, Daniel M. German, Margaret-Anne Storey, "Open source software peer review practices: a case study of the apache server", ICSE '08: Proceedings of the 30th international conference on Software engineering, May, pp541-550 (2008)
[10]   Dewayne   Perry,   Adam   Porter,   Michael Wade, "Reducing inspection interval in large-scale software development", IEEE Transactions on Software Engineering archive, Vol. 28, Issue 7, PP.695-705, July (2002)
[11] KOMIYAMA Toshihiro, "Development of Foundation for Effective and Efficient Software Process Improvement (<Special Features> Study of The Latest Trend of Software Management Technology)", IPSJ Magazine, Vol. 44, No4, Apr., pp341-347 (2003)

# Technology for Recommending Optimum Learning Texts Based on Data Mining of Learning Historical Data

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

[*] Department of Information Environment, Tokyo Denki University, Japan
[**] Graduate School of Information Environment, Tokyo Denki University, Japan
[***] Faculty of Software and Information Science, Iwate Prefectural University, Japan
yujiwada@sie.dendai.ac.jp, DEN03738@nifty.com , dohi@sie.dendai.ac.jp, and
sawamoto@iwate-pu.ac.jp

***Abstract*** - We are developing a bidirectional recommendation system that extracts the relationship among digital texts with historical logs, and recommends the optimum texts for learners using data mining methods, such as collaborative filtering. In this paper, we first discuss the bidirectional recommendation and then show results from an evaluation of actual use. Finally, we propose a method for a collaborative learning recommendation system that mines the data of similar users sharing non-favorite subjects using historical logs and user attribute data.[1]

***Keywords***: e-learning, data mining, recommendation, historical log analysis, collaborative filtering.

## 1 INTRODUCTION

In recent years, large numbers of institutions of higher learning, businesses and other organizations have been proactively introducing e-learning. That movement has been fostered in part by attention focused on the Web Based Training (WBT) approach [1], leading to the debut of numerous Learning Management Systems (LMS) [1]. Additionally, the proposal of the Sharable Content Object Reference Model, or SCORM [2], which is a global standard, has helped to spur the propagation of e-learning. Opinions are divided, however, as to whether the use of e-learning offers greater advantages to the learner than learning based on paper materials.

To address that question, firstly we implemented a "bidirectional recommendation system" [3] (see Figure 1) developed in our laboratory, in the AIRS "An Individual Reviewing System"[4]. Figure 1 shows a schematic of a bidirectional recommendation. When a learner is browsing the learning text "Basics of substitution," it is natural to advance to the next step "Basics of 'while' statement" or "Basics of 'if' statement." However, browsing the basic contents "Variable types" again is also natural in learning. In other words, the learning efficiency is expected to improve by recommending not only learning texts frequently shifted from but also frequently shifted to "Basics of substitution."
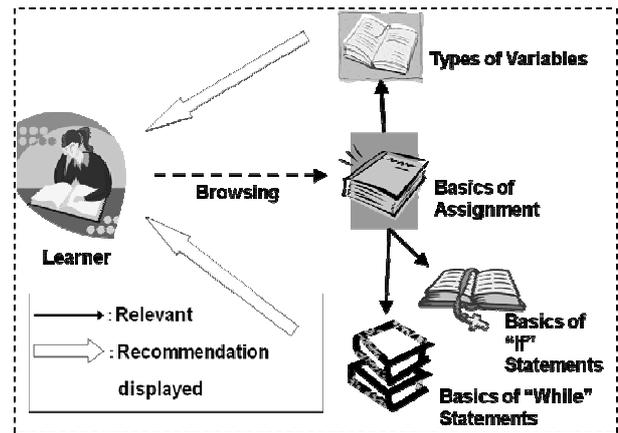
Figure 1: Bidirectional recommendation system.

Secondly, we asked 92 participants of the "Database System" lecture offered by the university to use the system between October 30 and November 5, 2008.

Subsequently, we conducted a survey using questionnaires that examined the actual situation of the user and the learning outcome achieved using the bidirectional recommendation system (see Table 1). In the survey, a number of respondents indicated that they were able to shorten the time spent learning, and the efficacy of learning using the bidirectional recommendation system was confirmed.

Table 1:User evaluation of bidirectional recommendation system.

| Opinions of those indicating that recommendation results were appropriate | Opinions of those indicating that recommendation results were not appropriate |
|---|---|
| It was easy to figure out what to look at next and less time was required. | I prefer to look at lecture materials and course handouts. |

Moreover, a recommendation accuracy of 61% resulted from subjective evaluation by users of the appropriateness of the recommendation results (see Table 2). Some respondents indicated, however, that they preferred to browse the lecture

materials, so we reexamined the functions requested by learners.

Table 2: Recommendation precision of bidirectional recommendation system.

| Percentage of respondents who said that the recommendation results were suitable, or somewhat suitable |
| --- |
| Recommendation precision = 61% |

We surmised that perhaps the objective of learners using this approach is to thoroughly review using the material used in lectures and deepen their understanding of it, even if it required more time. Based on this, we hypothesized the necessary function to be support information used when reviewing. For example, this could refer to "areas of weakness" that the learner finds harder to understand than the rest of the text.

Based on this, we proposed a "collaborative learning recommendation system" using e-learning, and developed a system designed to improve learning efficacy by recommending "areas of weakness" (hereafter referred to as "non-favorite subject material").

Finally, we surveyed the state-of-the-art about the recommendation technologies such as collaborative filtering and data mining, as follows. In [5], the design and implementation of a recommender system using social networks was described. In [6], a web content recommendation system based on the similarities is proposed. In [7], collaborative filtering based on C-SVM(Support Vector Machine) was proposed examined. In [8], data mining technologies, such as clustering and sequential pattern mining, for online collaborative learning data are studied. In [9], monitoring online tests, such as learner behavior and test quality, through data visualization are discussed. In [10], an automated learning and skills training system for a database programming environment is presented.

## 2 AIRS AND UTILIZATION STATUS

AIRS is an e-learning system that focuses specifically on review, and was developed starting from fiscal 2004 (see Figure 2).

Figure 2 displays the AIRS Japanese top page after a learner, who is going to review the database contents especially selection function, logs in AIRS. This page is comprised of the book marks (located at the upper side) of the contents available with AIRS, the contents menu (located at the left side) corresponding to the selected book mark, and the help messages for beginners (located at the right side). In figure 2, the contents menu displays database, data model, RDB, design methodology, and SQL. The learner selects the book mark such as selection and sorting before the learner can select the corresponding database contents menu. Then, the learner can proceed to review the database contents.
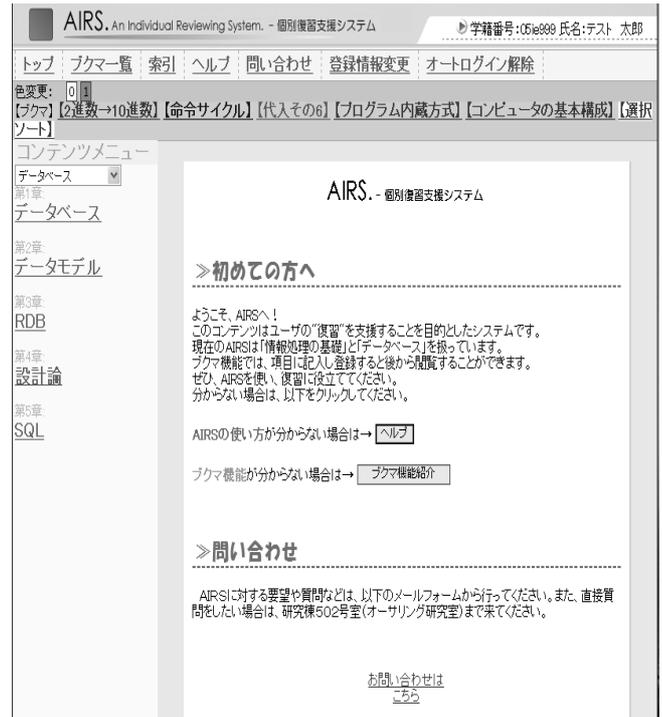


Figure 2: AIRS Japanese top page after login.

By focusing solely on review, the system reduces the possibility that the learner will rely on e-learning instead of sufficiently participating in lectures.

The system is also designed with the aim of improving learning efficacy through the synergistic effect of lecture-based learning and e-learning.



Figure 3: Example: contents of text.

One feature of the system is that it is an e-learning system by learners, for learners. This reason is that system is

designed to make learning easier, by having the developers attend lectures corresponding to the teaching content and develop content, to some extent, by anticipating sections that learners would have difficulty understanding (see Figure 3).

Additionally, the system is constructed so that each item being taught is expressed in three different ways (not yet fully implemented), and a function is provided by which learning is tailored to the individual learner, with the appropriate "form of expression" (refer to [11]) for that particular learner being automatically extracted.

The system comprises a database server that runs databases used by functions, such as the one mentioned previously, a content server that makes teaching content available, and a system server that runs AIRS (see Figure 4).
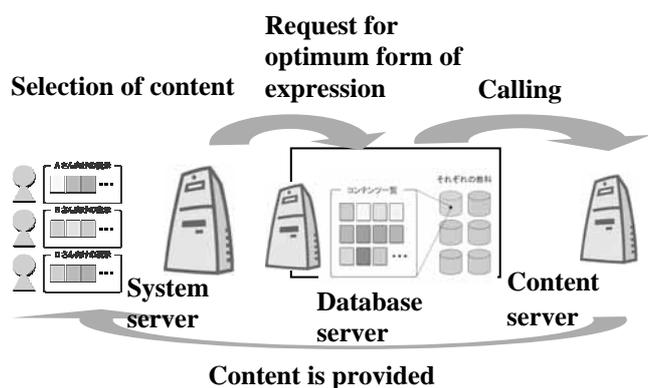


Figure 4: Configuration of AIRS.

The results of a questionnaire survey, conducted this fiscal year, concerning utilization status are shown in Figure 5. As previously mentioned, the survey targeted 92 participants of the "Database System" lecture offered by the university.
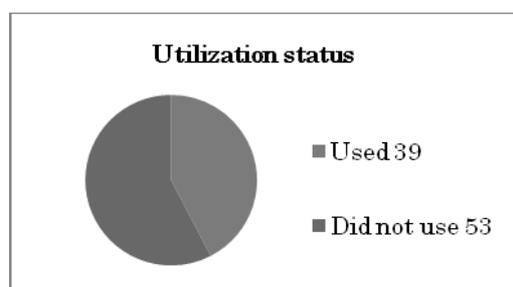


Figure 5: Usage achievements of AIRS.

# 3 BIDIRECTIONAL RECOMMENDATION SYSTEM

The aim of the system is to make it possible for learners to learn efficiently, without having to worry about selecting the content that was expanded through propagation of SCORM.

Moreover, based on the browsing history data of AIRS, it was found that an extremely large number of learners are sequentially browsing the course material in accordance with the flow of the material displayed on the screen. This is not different from review using paper materials and suggests learning efficacy will decrease as the volume of material expands. In actuality, the number of AIRS nodes (not taking the "form of expression" into consideration) has grown to

210 for two subjects. To solve these problems, the bidirectional recommendation system was developed as a means for recommending course material that is strongly relevant to the material currently being read, and thus improving both review efficiency and speed.

An overview of the system is presented here, together with a detailed description of the questionnaire previously described.

## 3.1 Overview

The system overview is presented in Figure 1. We assume here that the learner is browsing the material for the course called "Basics of Assignment" under "Fundamentals of Information Processing". At this point, the learner would naturally shift to the next steps, "The Basics of the 'While' Statement" and "The Basics of the 'If' Statement".

However, during the review process, it would not be unnatural for the learner to go back and re-read "Types of Variables", which is part of the basic content. In other words, learners could review the material, if the system, instead of recommending only material to which many learners shift after reading "Basics of Assignment" at the same time, recommends material that many learners read before moving to "Basics of Assignment". Looking back over material is a fundamental part of the review process, and we could expect an improvement in learning efficiency. This is why bidirectional recommendations are necessary, and is a feature of the bidirectional recommendation system.

## 3.2 Evaluation

Users of the bidirectional recommendation system filled out questionnaires regarding the number of times they used the system, the recommendation results, learning efficiency, whether or not they would like to use the system in the future, operability and other questions. The results are shown in Figure 6.

The targeted users and the organizations conducting the survey are the same as those for the questionnaire survey previously described.

The results indicated a large number of learners used the system infrequently because they had problems logging in. Even taking that into consideration, the majority of users obtained favorable results using the system for only one week, and the system can be expected to improve learning efficacy.

Firstly, 64% of the respondents said that the system was useful as shown in Figure 6 (b).

Secondly, 66% of the respondents said that the recommendation results were suitable as shown in Figure 6 (c).

Thirdly, 58% of the respondents said that efficiency improved as shown in Figure 6 (e).

Moreover, 64% of the respondents said that they would use the system in the future, including those who thought the contents were easy to understand and those who would use it if errors were corrected, as shown in Figure 6 (f).

However, 74% of the respondents said that the recommendation function is not easy to operate as shown in Figure 6 (g).
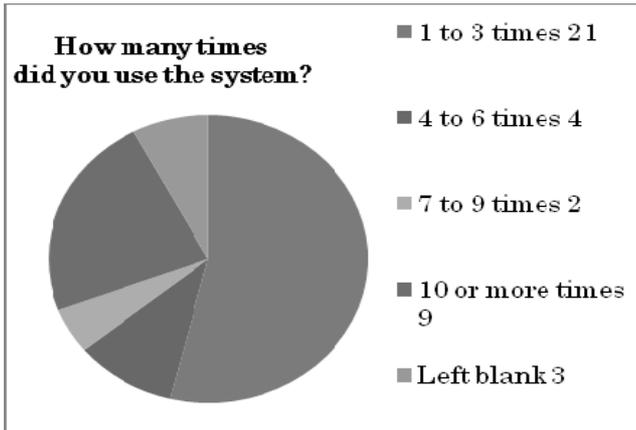
(a) What is the frequency of use?
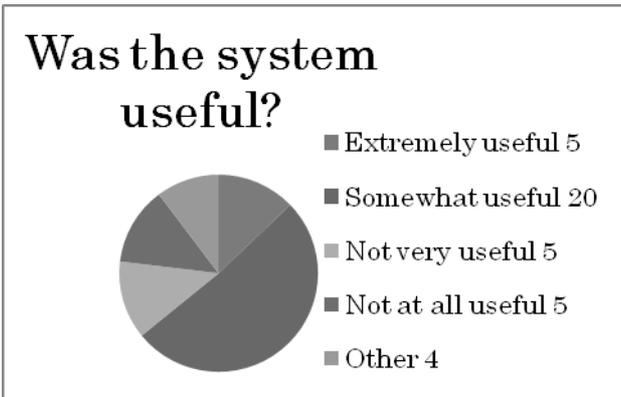


(b) Is it useful?
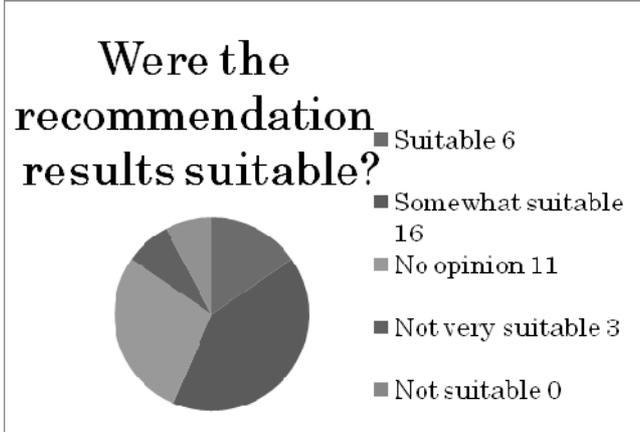


(c) Is it suitable?
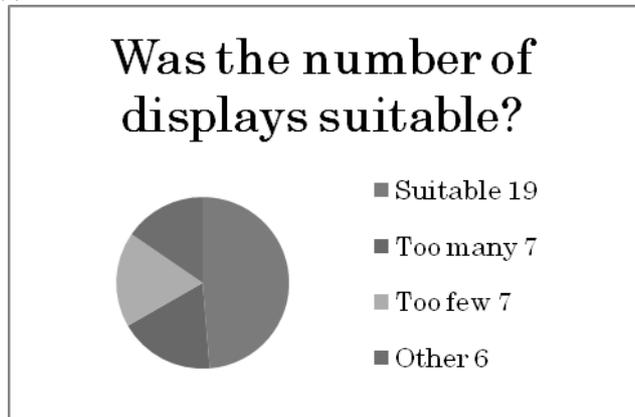


(d) What is the number of displays?
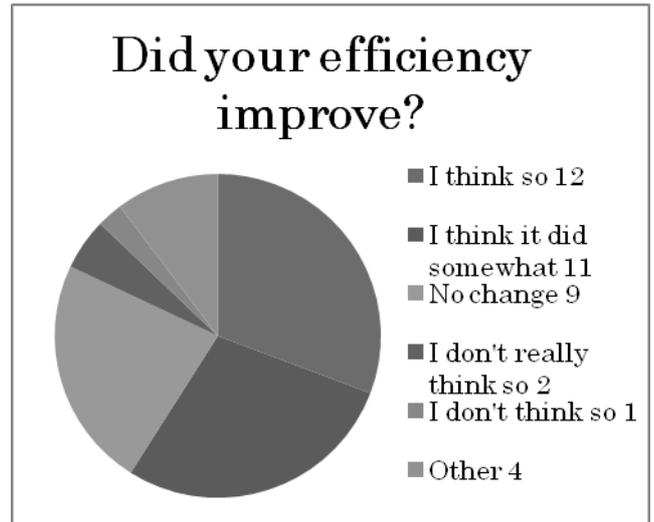


(e) Is it efficient?
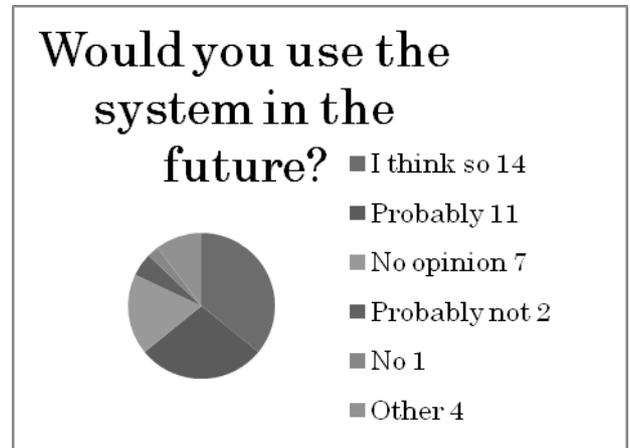


(f) Will you use it in the future?
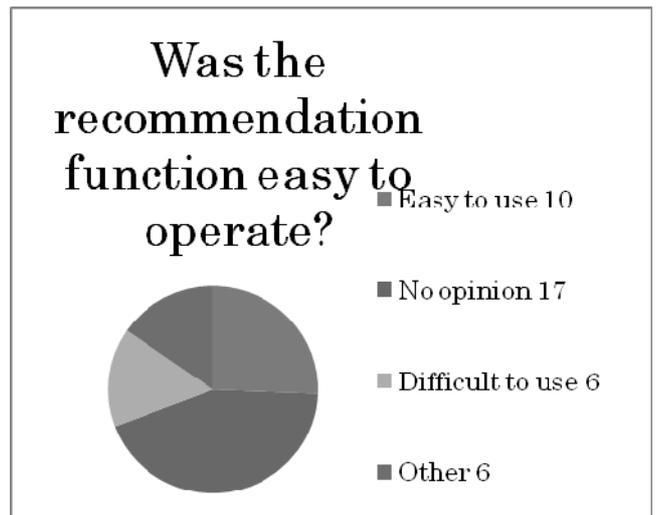


(g) Is it easy to operate?

Figure 6: User evaluation results of bidirectional recommendation system.

# 4 COLLABORATIVE LEARNING RECOMMENDATION SYSTEM

For the learning history data accumulated in AIRS, we focused on the number of times that the learners had browsed the various subject materials, suspecting that frequent browsing indicated that the material had not been sufficiently understood from the lecture, and that there was a strong possibility of this being an area of weakness for the learner.

However, if subject material that the learner had browsed numerous times was recommended as "non-favorite subject material", without any additional input or modification, there was a strong possibility that the learner was already aware of this material, having browsed it multiple times, and that this would not improve learning efficacy. There were also concerns that the recommendations would be concentrated too heavily on the same subject material, creating a convergence in the recommendation results.

To resolve this, we focused on collaborative filtering technology [12] in our research. Using this technology, it was possible to identify similar users with similar browsing histories from among the learning history data (see Figure 7).

It was thought that by recommending frequently browsed material from the learning history data of similar users, it would be possible to provide "non-favorite subject material" that the learner did not yet know (see Figure 8).
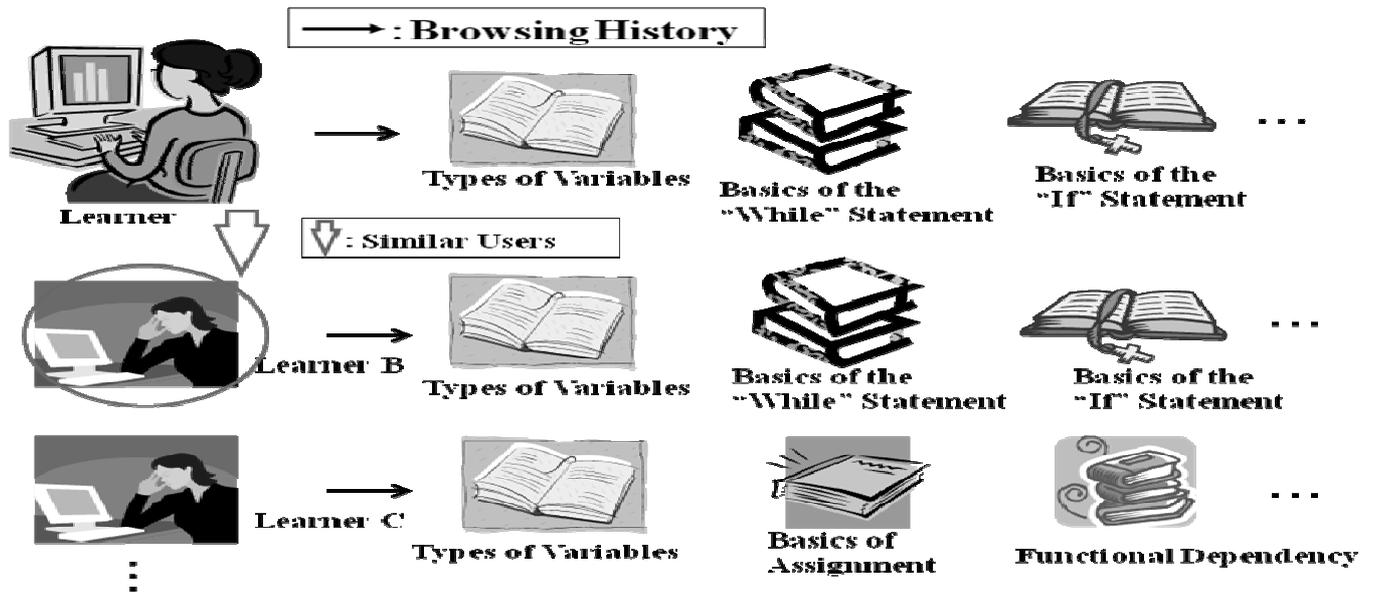


Figure 7: Mining of browsing history data to identify similar users
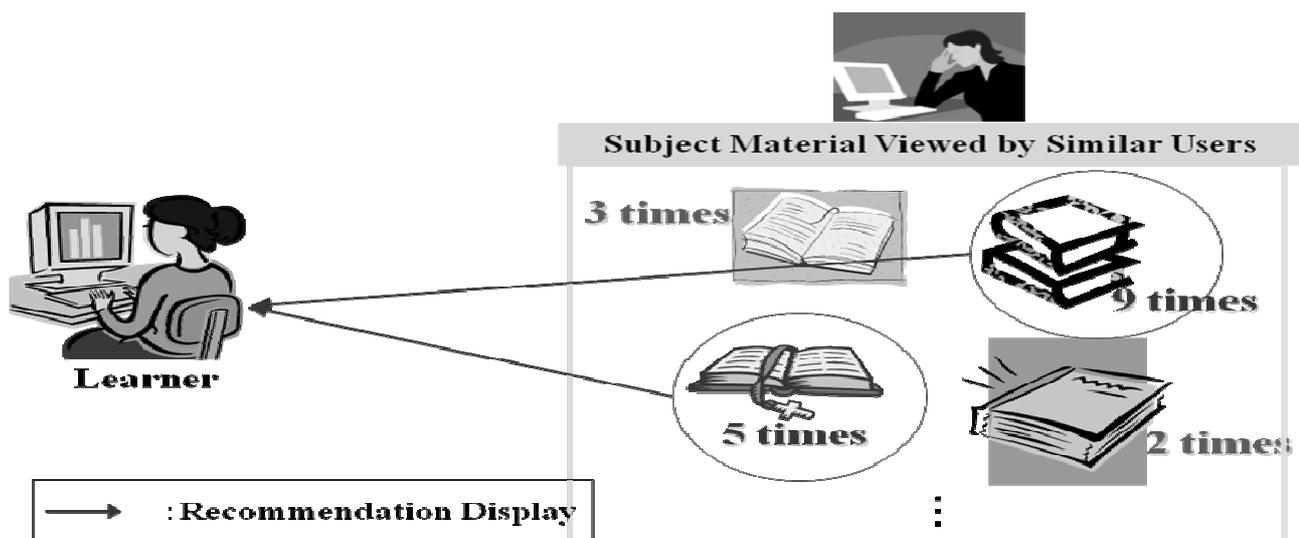


Figure 8: Data mining of non-favorite subject material.

# 5 MINING DATA OF SIMILAR USERS

## 5.1 Attribute Data and Similar Users

Information not obtained by AIRS includes the learner attribute data, such as age, gender, hobbies and preferences.

In the present study, the strong subjects, non-favorite subjects, average learning time, hobbies and preferences, number of AIRS logins, usage time and other parameters of the user were additionally defined as learner attribute data. The purpose of acquiring this attribute data was to provide detailed recommendations even if the user was new to the system.

When browsing histories of similar users are mined (see Figure 7), new users are unable to find similar users because they have no learning history data, and recommendation accuracy drops sharply as a result. When all users have the same attribute data, it becomes possible to mine data for new users and similar users as well. The method for mining attribute data of similar users is shown in Figure 9.

The collaborative filtering method was used for mining the data of similar users, and mining of non-favorite subject material was done as described in section 4 (see Figure 8).
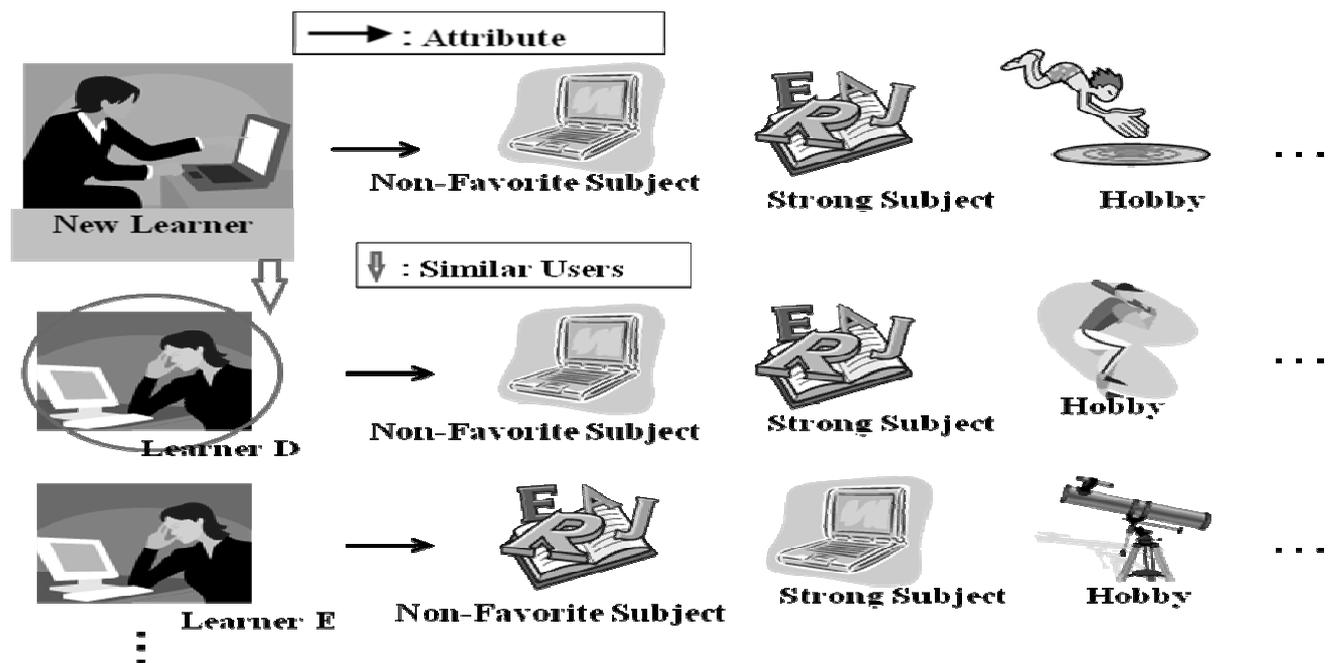


Figure 9: Method for mining attribute data of similar users.

## 5.2 Effective Attribute Data Group

In the present study, mining all of the attribute data would not be useful in identifying similar users. Therefore, it was considered important to identify "attribute data groups" that were useful or effective, consisting of combinations of several attribute data elements.

A method proposed for identifying these attribute data groups is shown in Figure 10.

(Step 1) First, one attribute data combination is created. For the time being, this is called the "first attribute group".

(Step 2) Similar users are identified, referring to this first attribute group (see Figure 9). These are "similar users based on the first attribute group".

(Step 3) The "similar users based on the first attribute group" identified at step 2 are compared to the "similar users based on browsing histories" identified using the method described in section 4 (see Figure 7), and the percentage of matches is calculated as the match rate.

(Step 4) The learner who will serve as the reference is substituted for Learner B, and the match rate is calculated by repeating steps 2 and 3. In the same way, the match rates for subsequent learners (e.g., Learner C, Learner D) are determined until match rates have been determined for all of the users. The total of the match rates for all users is then divided by the number of users (n) to find the mean match rate, and that value is used as the "effective index of the first attribute group".

(Step 5) An attribute data group is created at step 1 again, and the process through step 4 is repeated.

This process is repeated and the effective index sequentially increased to find the "most effective attribute data group".
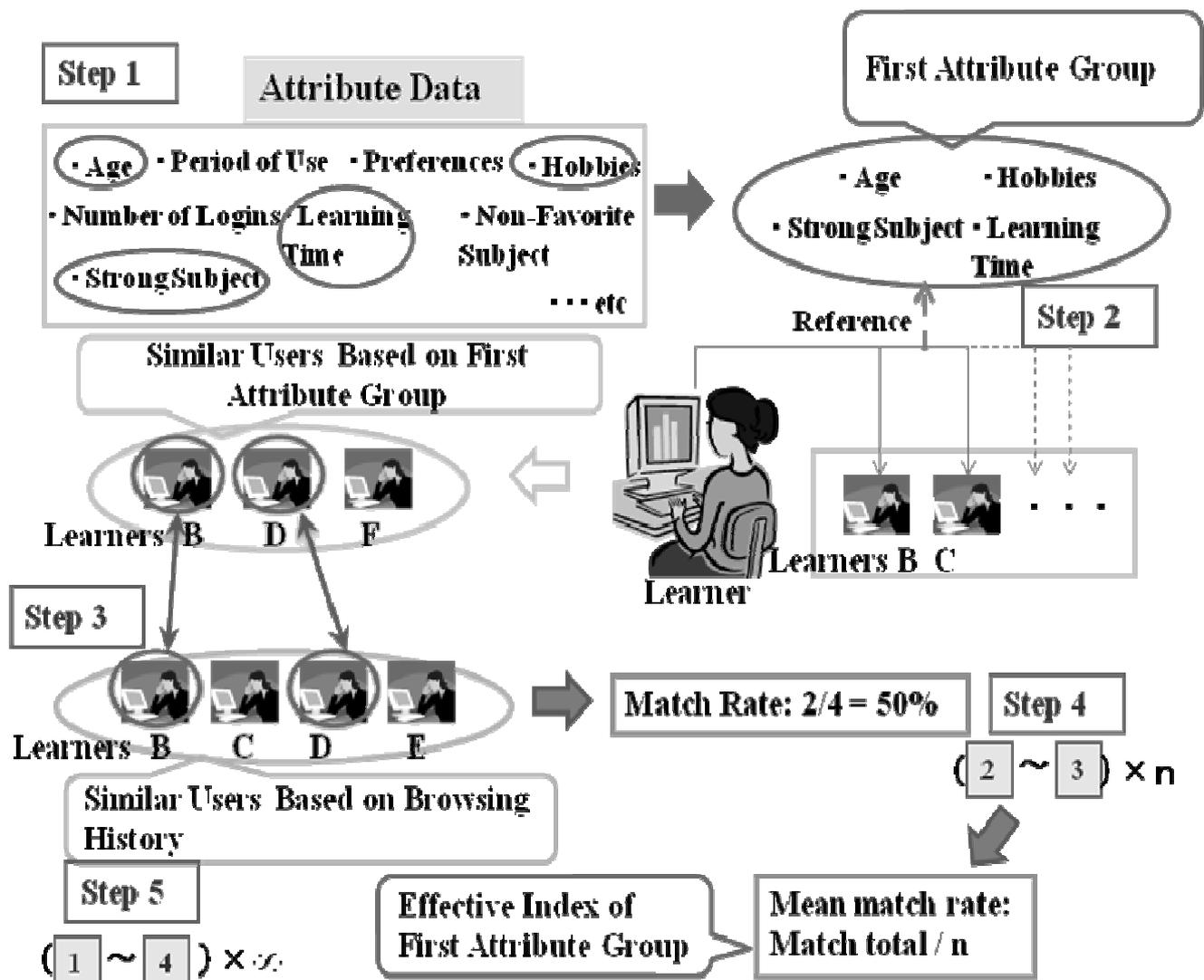
Figure 10: Method for mining useful attribute data.

## 6 CONCLUSION

In this paper, we proposed a method for a collaborative learning recommendation system that mines the data of similar users sharing non-favorite subjects using historical logs and user attribute data.

The method for mining non-favorite subject material proposed here is based on the assumption that the more times the content has been browsed, the less skilled the learner is in that subject.

For this reason, we currently plan to develop a collaborative learning recommendation system and implement it in the AIRS, and to verify the appropriateness of the recommendation results by measuring recommendation precision.

Then, the recommendation precision will be measured using the following data:

(1) Questionnaire results reflecting the subjective view of the student (user),

(2) Information relating to teaching material in which the learner is thought to be weak (as indicated by the course instructor),

(3) Comparison results of learning effectiveness between students who used AIRS with collaborative learning recommendations and students who used AIRS without these recommendations.

Finally, we consider our future work is as follows: we collect new attribute data, we ascertain the usefulness and effectiveness of the attribute data, and we evaluate the recommendation results for new users.

## ACKNOWLEDGMENTS

# REFERENCES

[1] Ministry of Economy, Trade and Industry, e-Learning White paper 2007/2008, Tokyo Denki University Press (2007).

[2] R. Wisher, Sharable Content Object Reference Model (SCORM) 2004 4th Edition Overview Version1.0, Advanced Distributed Learning, pp.1-18 (31 March 2009).

[3] Y.Wada, T.Matsuzawa, M.Yamaguchi, and S.Dohi, Proposal and Verification of Bidirectional Recommendation System for Learning Web Digital Texts, Proceedings of the Second International Conference on the Applications of Digital Information and Web Technologies (ICADIWT2009), pp.210-215 (August 2009).

[4] M.Yamaguchi, Y.Takahashi, T.Matsuzawa, T.Takahashi, and S.Dohi, Development and Evaluation of e-Learning System AIRS with Multi-Expression Contents, IEICE-ET2006-118, Vol.106, No.583, pp.69-74 (2007).

[5] H.Hotta, and M.Hagiwara, Design and Implementation of Social-network-based Recommender System, IPSJ Transactions on Databases, Vol.2, No.1, pp.46-56 (2009).

[6] A.Sasaki, T.Miyata, Y.Inazumi, A.Kobayashi, and Y.Sakai, Web Content Recommendation System Based on Similarities among Contents Cluster of Social Bookmark, IPSJ Transactions on Databases, Vol.48, No.SIG20, pp.14-27 (2007).

[7] K.Oku, S.Nakajima, J.Miyazaki, and S.Uemura, Context-Aware SVM for Context-Dependent Information Recommendation, DBSJ Letters, Vol.5, No.1, pp.5-8(2006).

[8]D. Perea, J. Kay, I.Koprinska, K.Yacef, and O.R.Zaiane, Clustering and Sequential Pattern Mining of Online Collaborative Learning Data, IEEE Transactions on Knowledge and Data Engineering, Vol.21, No.6, pp.759-772(2009).

[9]G.Costagliola, V.Fuccela, M.Giordano, and G.Polese, Monitoring Online Tests through Data Visualization, IEEE Transactions on Knowledge and Data Engineering, Vol.21, No.6, pp.773-784(2009).

[10] C.Pahl and C.Kenny, Interactive Correction and Recommendation for Computer Language and Training, IEEE Transactions on Knowledge and Data Engineering, Vol.21, No.6, pp.854-865(2009).

[11] Y.Takahashi, T.Matsuzawa, M.Yamaguchi, S.Dohi, and Y.Wada, Recommendation and Delivery of the Optimum Texts for e-Learners, IPSJ SIG Technical Reports 2007-CE-88(22), Vol.207, No.12, pp.157-162 (2007).

[12] N.Yamazaki, Software Design ISSN 0916-629, Vol... 268 Gijutsu-Hyohron Co., Ltd, Tokyo Japan(2007-8).

# Session 4: Networking
## (Chairs Teruo Higashino and Tomoya Kitani)

# Proposal and Implementation of Coordinate Integrations for Heterogeneous Network Protocols

Tomoya Takenaka[†], Hiroshi Mineno[‡], and Tadanori Mizuno[†]

[†]Graduate School of Science and Technology, Shizuoka University, Japan
[‡]Faculty of Informatics, Shizuoka University, Japan
{tomoya, mineno, mizuno}@mizulab.net

***Abstract*** - Location estimation is important for ad-hoc and sensor networks. Existing localization techniques assume to be operated in a single network protocols. We propose coordinate integrations for heterogeneous network protocols. The fundamental concept underlying coordinate integrations is that sets of coordinates is iteratively integrated by using at least three shared nodes in two-dimensional space so that coordinates generated in different network protocols become compatible across the networks. We used a simulation to demonstrate the proposed coordinate integration. We also present details of implementations on sensor nodes and experimental results for RSSI measurements inside a university building.

***Keywords***: Localization, location estimation, sensor networks, heterogeneous networks, ad-hoc networking

## 1 Introduction

Ad-hoc networking enables a wireless network to be constructed without an infrastructured base station. Ad-hoc networking also enables nodes to relay a data using multi-hopping. Emerging products of sensor networks such as Zigbee use ad-hoc networking capabilities over IEEE 802.15.4 [6]. Other wireless networking technologies such as wireless local area Networks (WLANs) and bluetooth support ad-hoc networking over IEEE 802.11 [4] and 802.15.1 [5].

Localization is an attractive functionality of using ad-hoc networking capabilities, which enables nodes to estimate their positions. The motivation for developing a localization technique is to inform an observer of the many deployed node positions with a small number of anchor nodes whose positions are known in advance. Location information is not only used for bundling sensing events with their locations, but also for improving network performance.

Anchor-free localization was proposed in [11], and it has received much attention. The advantage of the anchor-free localization is that it enables nodes to estimate their positions without using anchor nodes. The set of node coordinates is relatively determined, and hence it assigns an arbitrary relative coordinate system in each network protocol.

Assume that the nodes are deployed over a field and that they have incompatible network protocol. When anchor-free localizations are applied to heterogeneous network protocols that have different network protocols coexisting, one coordinate system is incompatible with other coordinate systems. This is because the set of node coordinates is relatively determined. Therefore, nodes cannot use location information

obtained using anchor-free localizations across other network protocols.

In this paper, we propose coordinate integrations for heterogeneous network protocols. Sets of estimated coordinates in heterogeneous network protocols are iteratively integrated using nodes that physically share same coordinates on different network protocols. The coordinates are then compatible for heterogeneous network protocols. We first conducted simulation evaluation to verify the proposed coordinate integration. We are currently implementing functionalities of proposed coordinate integration on sensor nodes, and present details of the implementation.

This paper is organized as follows. Related work and localization issue in heterogeneous network protocols are described in Section 2. Proposed coordinate integration is presented in Section 3. An evaluation of coordinate integration using simulation is presented in Section 4. A detail of our implementations of coordinate integration is presented in Section 5. Results for RSSI experiments are reported in Section 6. Section 7 concludes the paper and mentions future work.

## 2 Related work and issue

### 2.1 Related work

Localization techniques have been discussed for wireless multi-hop networks such as sensor and ad-hoc networks. The motivation behind developing multi-hop localization is wanting to know where the node position is in wireless multi-hop networks by using a small fraction of the anchor nodes. An anchor node is one whose position is known in advance through means such as global positioning system (GPS). A simple solution to obtaining location information is to equip each node with a GPS receiver. However, a GPS receiver cannot always receive signals from GPS satellites when it is located in a building, and it enforces equipment costs for nodes. Much research has been conducted on how to estimate node positions in wireless multi-hop networks. Most localization techniques can be categorized into two types. The first is localization by using extra ranging devices, such as ultra sound devices, and the second is localization without using extra ranging devices.

AHLoS [10] is the distance-measurement localization approach using ultra-sound ranging devices. In AHLoS, at least three anchor nodes iteratively conduct multilateration to estimate unknown node positions. Once the positions for unknown nodes are estimated by anchor nodes, the nodes are
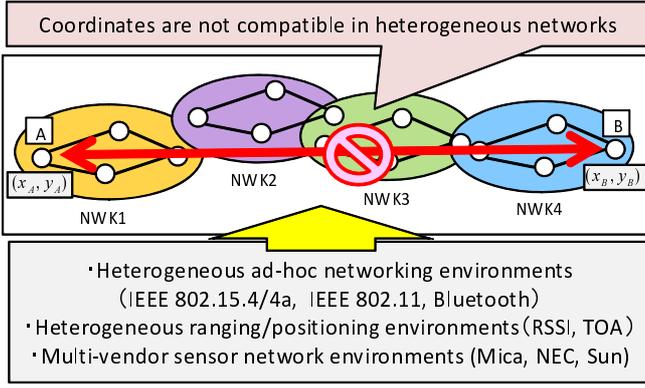
Figure 1: Localization issue in heterogeneous network environments.



Figure 2: Conceptual illustration of proposed coordinate integrations.

configured as pseudo-anchor nodes and estimate unknown nodes that remain in the network. In sweeps [13], algorithms to identify global rigidity were employed to estimate the node positions without flipping for sparse node networks. The distance-measurement approach normally achieves precise positioning accuracy. However it requires extra ranging devices, increasing the cost for all nodes.

The localization scheme without using extra ranging devices has been developed for large-scale sensor networks, and it exploits connectivity information of multi-hop networks. In DV-Hop [9], the positions for unknown nodes in a network are estimated by using trilateration using average hop-count distances from at least three anchor nodes. In anchor-free localization (AFL) [11], the positions of unknown nodes are estimated without using anchor nodes. The basic idea behind AFL is to select reference nodes that represents the relative axis in a network and they determine relative node positions based on the hop-counts from their reference positions..

Some research proposed to use multidimensional scaling (MDS) [12, 14] to estimate node positions. MDS is the statistical technique to obtain geographical representations of data from data proximity. Since distance information can be used as data proximity, MDS can plot the relative coordinates of nodes in a network. Basic idea to apply MDS to localization technique is that node collects distance information such as hop-count and TOA measurement and calculates relative node positions by using MDS.

The localization scheme without using ranging devices enables nodes to estimate node positions while only using the radio capabilities of a sensor node. Hence, it has great flexibility to enable nodes to be applied to localization in the network. However, existing localization techniques are assumed to be operated in single network protocols.

## 2.2 Localization issue in heterogeneous network environments

Figure 1 presents a localization issue in heterogeneous network environments. Some localization techniques require anchor nodes that have the unique original point of the coordinate system. For example, GPS has the original point that is centroid of earth in the coordinate system. The coordinate system has the original points in its own network, how-

ever these coordinate system can be arbitrary determined in the network. In addition, in anchor-free localization, the estimated coordinates are relatively determined and original points are arbitrary determined. Therefore, coordinate systems that are generated by using localization techniques in different network protocols are not compatible with each other. Especially, localization technique using ad-hoc networking capability can be more general technique for position estimations since anchor-free localization only requires ad-hoc networking capability to estimate node positions. Currently, several IEEE standardized networking protocols such as IEEE 802.11 [4]  IEEE 802.15.4 [6]  IEEE 802.15.1 [5] (bluetooth) supports ad-hoc networking capability. However these networking protocols are not able to communicate with different network protocols. In the sensor networks, networking protocols such as Zigbee [1] have been standardized. Several vendors such as MICA and NEC have released sensor nodes in market. However, one vendor's sensor node cannot communicate with another vendor's sensor node. In heterogeneous network protocol environments, we encounter an issue that one set of coordinates generated by a network cannot be used in another network as shown in Fig. 1. Hence, a mechanism that converts coordinates generated in heterogeneous network into one set of coordinates that is compatible across the networks is required.

## 3 Coordinate integrations for heterogeneous network protocols

### 3.1 Overview

Figure 2 shows a conceptual illustration of proposed coordinate integrations. We consider that heterogeneous network protocols are coexisted in a field as shown in bottom of Fig. 2. Each network protocol conducts localization techniques by using ad-hoc networking capability. In the coordinate integration, coordinates estimated by each localization protocols are integrated by using at least three shared nodes as shown in middle of Fig. 2. The shared node is the node that physically
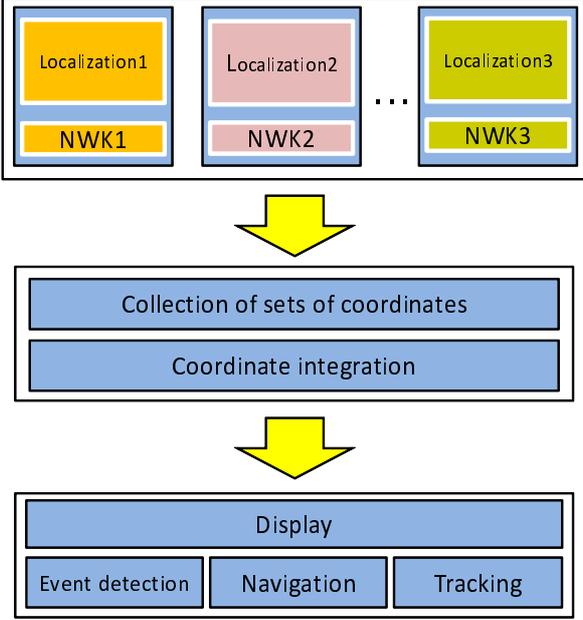
Figure 3: Logical organization for proposed coordinate integration.

shares same coordinates in different network protocols. Some nodes in different network protocols are set as shared nodes for the coordinate integration. The integrated coordinates are compatible with different network protocols and they can be accessed by users as shown in top of Fig. 2.

Coordinate integrations for heterogeneous network protocols are conducted as follows.

1. **Coordinate assignments:** Each node in the heterogeneous network protocols estimates its relative or absolute node coordinates within its network protocol. Developing the precise localization algorithm is not our focus for this paper. Reader may refer to the literature [11].

2. **Coordinate conversions:** The orientations of two sets of coordinates are adjusted into one set of coordinates by using three or more shared nodes. The coordinate conversion requires rotation, flipping and translation operations. Here, we used procrustes analysis [16] for the coordinate conversions.

3. **Coordinate integrations:** Two sets of coordinates in different network protocols are integrated into one set of coordinates. The other sets of coordinates are iteratively converted and integrated into one set of coordinates. The coordinate system is then compatible for heterogeneous network protocols.

Figure 3 presents the logical organization of proposed coordinate integration. When nodes in heterogeneous network conduct a localization technique, one set of coordinates of network that is incompatible is generated (top of Fig. 3). One localization server collects all sets of coordinates of networks, and conduct coordinate integration to be compatible with each other (middle of Fig. 3). The coordinates that is compatible

with heterogeneous network protocols are available for application such as event detection, tracking and navigation regardless of the network protocols (bottom of Fig. 3).

## 3.2 Algorithm

The objective of proposed coordinate integration is to convert sets of coordinates that are incompatible with other coordinates of networks into a set of coordinates that is compatible with other coordinates of networks. Therefore, localization technique to obtain estimated coordinates is not our main focus of the proposed coordinate integration. In this work, we assumed to use MDS that can obtain the relative coordinates of networks from distance measurements such as RSSI.

MDS [16] is statistical technique used to analyze proximity data in multidimensional space. The proximity data for MDS can be represented by geographical expression. Therefore proximity matrix that is constructed by using node distances can be transformed to the relative coordinate system by using MD.

MDS to obtain relative coordinates from node distance is operated as follows. First a node constructs the squared distance matrix,

$$\mathbf{D}^{(2)} = \{\hat{d}^2_{\{k,l\}}\}, \tag{1}$$

where the number of nodes is $n$ and the node distance between $(i, j)$ is denoted by $d_{ij}$. Multi-hop distances are approximated by using hop-couting. The scalar product matrix, $\mathbf{B}$ is constructed by applying double centering as

$$\mathbf{B} = -\frac{1}{2}\mathbf{J}\mathbf{D}^{(2)}\mathbf{J}, \tag{2}$$

where $\mathbf{J} = \mathbf{I}_n - \frac{1}{n}\underline{1}\underline{1}^T$ and $\underline{1}$ is an $n$ by 1 vector of ones and $n$ is the length of $\mathbf{D}_i$. A singular value decomposition is conducted as

$$\mathbf{B} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T. \tag{3}$$

A coordinate matrix is then given by $\mathbf{X} = \mathbf{U}\boldsymbol{\Lambda}^{1/2}$. Node position $P_i^0(x_i^0, y_i^0)$ is obtained by extracting the first and second columns of $\mathbf{X}_i$.

When the coordinate system is relative such as MDS or the original points of coordinate systems are different, coordinate systems are not compatible with each other. In order to address this issue, we used procructes analysis [16] to rotate sets of coordinates. Procrustes analysis for coordinate rotation is operated as follows. Consider that two sets of coordinates $X, Y$ that are not compatible with each other. Each set of coordinates is represented by matrix. Two sets of coordinates assume to have at least three shared coordinates. Objective of rotation operation of coordinates is to derive rotation matrix $\mathbf{T}$. First two sets of coordinates are multiplied as $\mathbf{A} = \mathbf{X}^T\mathbf{Y}$. We then calculate singular value decomposition for matrix $\mathbf{A}$ as $\mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{M}^T$. The rotation matrix for coordinates is then derived as

$$\mathbf{T} = \mathbf{M}\mathbf{L}^T. \tag{4}$$

The coordinates $Y$ can be compatible with coordinates $X$ by

$$\mathbf{Z} = \mathbf{T}^T\mathbf{Y} + c, \tag{5}$$

(a) Min neighbor    (b) Random    (c) GDOP-based
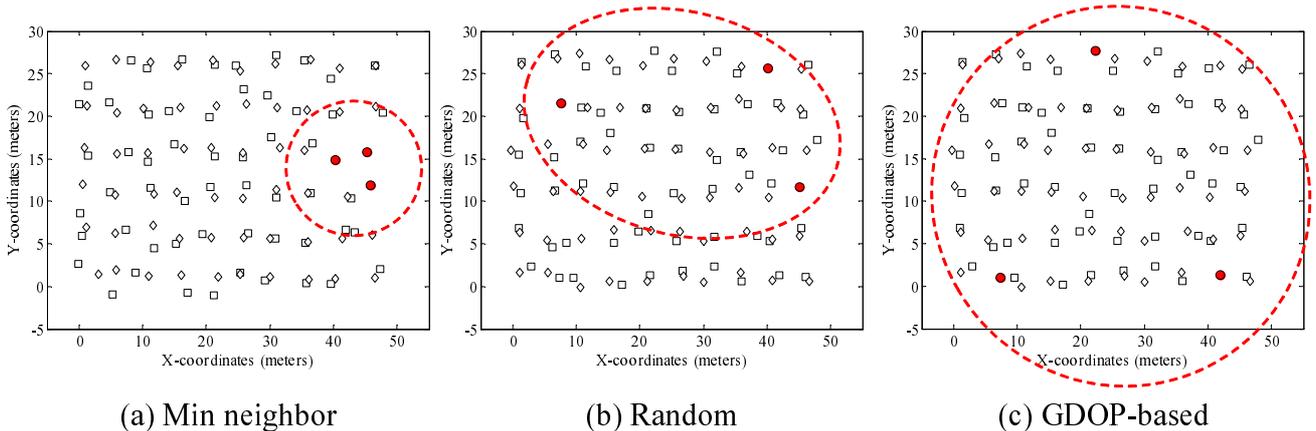
Figure 4: Shared node placements for different selection methods. (a): Min neighbor method, (b): random method, (c): GDOP-based method. Nodes for two different network protocol represented by squares and diamonds are overlapped. Shaded circles indicate shared nodes.

where $c$ is translation component. Finally, in order to derive the coordinates $\mathbf{R}$ that is compatible with $X, Y$, two set of coordinates are integrated into one set of coordinates by averaging as

$$\mathbf{R} = \frac{1}{2} * (\mathbf{X} + \mathbf{Z}). \tag{6}$$

Sets of coordinates are repeatedly integrated for other coordinates of networks based on shared nodes.

### 3.3 Placement for shared nodes

#### 3.3.1 Min neighbor method

In the coordinate integration, at least three shared nodes that have same coordinates for different networks are required. Here we discuss three methods on placements for shared nodes.

Figure 4 shows three patterns of node placements for shared nodes. In Fig. 4, nodes for two different network protocol represented by squares and diamonds are overlapped and co-existed in a field. Shaded circles indicate shared nodes.

One possible method for determining shared node placement is min neighbor. In min neighbor methods, one node is randomly selected and other two nodes that are closest to it are selected as shown in Fig. 4(a). The advantage of the method is that it can be easy to place the share nodes for heterogeneous network environments.

#### 3.3.2 Random method

Random method is that nodes for shared coordinates are randomly selected as shown in as shown in Fig. 4(b). The random method suggests that user does not take care of location for share nodes.

#### 3.3.3 GDOP-based method

GDOP-based method selects shared nodes based on the geometrical dilution of precision (GDOP) [8]. GDOP was used

Table 1: Simulation parameters.

| Parameter | Value |
|---|---|
| $n_p$ | 2.5 |
| $P_t$ | $-50.0 - -35$ (dBm) |
| $\sigma_{dB}^2$ | 5.0(dB) |
| Receiver sensitivity | $-73.0$ (dBm) |

to indicate the geometric conditions of the anchor nodes (i.e., GPS satellites in GPS) defined as the following equation.

$$GDOP = \sqrt{\frac{N_A}{\sum_{i \in S_A} \sum_{j \in S_A, j > i} A_{ij}^2}}, \tag{7}$$
$$A_{ij} = \sin \theta_{ij}.$$

In Formula (7), $N_A$ indicates the number of anchor nodes, and $S_A$ is the set of anchor nodes, and $A_{ij}$ is the angle from an unknown node to anchor nodes $\{i, j\}$. When GDOP is small value, condition for anchor node to estimate unknown node would be good. GDOP takes small value when the area of anchor nodes is larger as shown in Fig. 4(c). GDOP-based method selects three nodes that have minimum GDOP when all nodes are assumed to be anchor nodes.

## 4 Simulation

### 4.1 Simulation setting

We tested the coordinate integrations using a simulation. The objective of using the simulation is to demonstrate the coordinate integrations for heterogeneous network protocols before the implementation. Simulation tool we used was Matlab. We assumed to use the received signal strength (RSS) to measure the distance. The RSS measurement can be modeled as [15]

$$P_r = P_t - 10n_p \log_{10}(\frac{d}{d_0}) + X_\sigma. \tag{8}$$

$\overline{P_0}$ (dBm) is the mean signal strength and $P_t$ (dBm) is the received signal strength at reference distance $d_0$. $X_\sigma$ is zero-
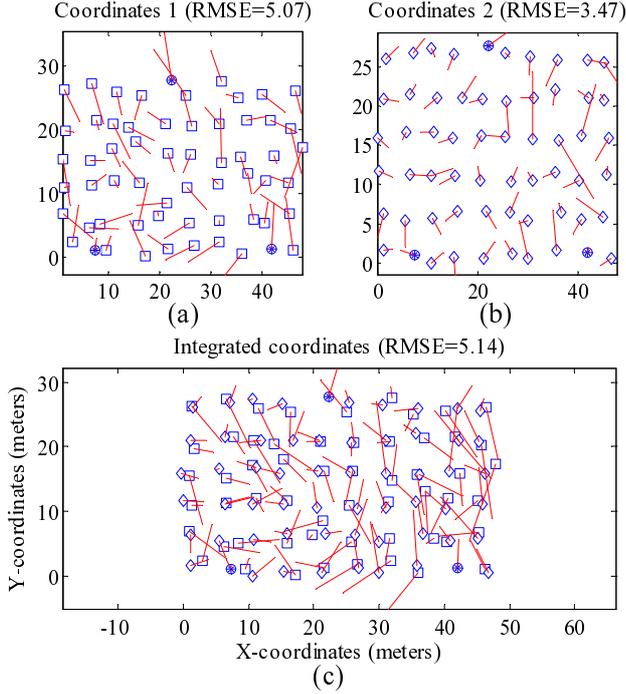
Figure 5: (a, b): relative coordinates systems for network 1 and 2. (c): relative coordinates system after coordinate integration. Actual node positions on network 1, 2 are represented by squares and diamonds, and their shared nodes are represented by asterisks. Errors are drawn by solid lines

mean Gaussian distribution with variance $\sigma_{dB}^2$ for the lognormal shadowing, and $n_p$ is path loss exponent determined in the measurement environment [15]. Table 1 shows the parameters used in the simulation. $n_p = 2.5$ and $\sigma_{dB}^2 = 5$ were chosen in the simulation. Field size of node placement is 30 (height) $\times$ 50 (width) $[m^2]$. We assumed two vender sensor networks that are incompatible with each other are deployed in the filed. The number of nodes on each network was 60.

## 4.2  Results

Figures 5(a)(b) show each set of estimated positions using the MDS, and Fig. 5(c) shows the integrated coordinates of two networks. The shared nodes are selected by using GDOP-based methods. The two sets of estimated coordinates in different network protocols were successfully integrated into one set of coordinates based on coordinates of shared nodes by using procrustes analysis. We defined root mean square error (RMSE) as follows

$$\mathbf{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\{(X_i - X_i^t)^2 + (Y_i - Y_i^t)^2\}}, \qquad (9)$$

where $(X_i, Y_i), i = 1\ldots N$ is estimated node position and $(X_i^t, Y_i^t)$ is actual node position. RMSE of coordinates 1, 2 and integrated coordinates were 5.0, 3.4, and 5.1 (m), respectively.

Figure 6 shows average RMSE plotted with varying node connectivity. Node connectivity shows how many nodes con-
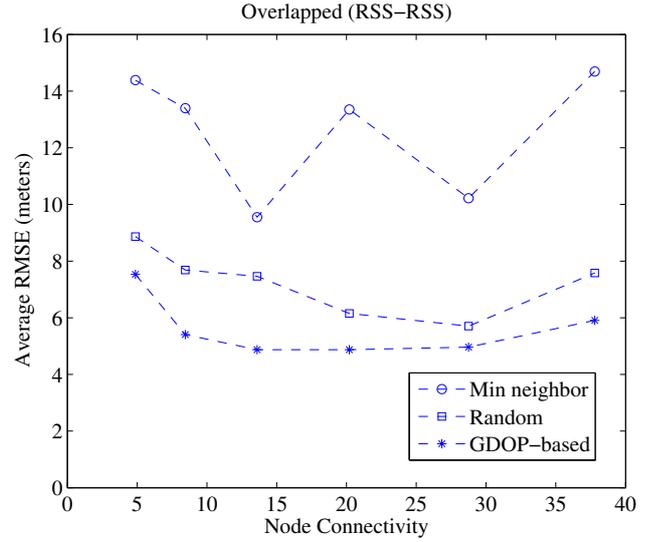


Figure 6: Average RMSE against node connectivity.

nect to other nodes in 1-hop on average, and it is varied by increasing communication range, i.e., increasing $P_t$. As shown in Fig. 6, GDOP-based method achieved best performance of other methods. GDOP-based method selects the three shared nodes that has larger regular triangles. This avoids flipping that has the wrong orientation of integrating two coordinate. On the other hand, min neighbor method had poor performance. This is because the integrated coordinates of two networks includes flipping results. Our simulation results only cover simple scenarios of integrating two coordinate systems. We are currently planning to conduct simulations with more complex scenarios including a mixed TOA/RSSI and different localization algorithms such as trilateration in large-scale heterogeneous networks. The detailed characteristics for coordinate integrations will be analyzed in the simulations.

As shown in Fig. 6, GDOP-based method increased the RMSE when the node connectivity was over 20. One explanation for this result is that attenuation model of RSS used in Formula (8) gets slower and variances get larger when the communication range is longer. Hence, when the communication range is increased, RMSE gets wrong. The result suggests that nodes are required to vary communication range to achieve less RMSE.

We discuss potential applications of using location information provided by the localization technique in sensor networks. Supporting location information in sensor networks enables us to develop many kinds of location-based applications such as firefighter navigation and equipment monitoring [7]. Our target application of localization technique is a smart air conditioning system using sensor networks [17]. Existing air conditioning system only utilizes single sensor such as infra-red (IR) to measure temperature in a room and the sensing coverage is limited. The smart air conditioning system called "i-fan" uses sensor networks to measure temperature and the corresponding location by using localization techniques. Although it is future work whether i-fan satisfies the location accuracy provided by our localization system, we are aiming to develop such location-based application.

Table 2: Basic specifications for Renesas and SunSPOT sensor nodes.

|  | Renesas node | SunSPOT |
|---|---|---|
| RF module | Freescale MC13202 | TI CC2420 |
| PHY/MAC | IEEE 802.15.4 | IEEE 802.15.4 |
| Frequency | 2.405–2.48GHz | 2.40–2.4835GHz |
| Receiver sensitivity | $-92$dBm | $-95$dBm |
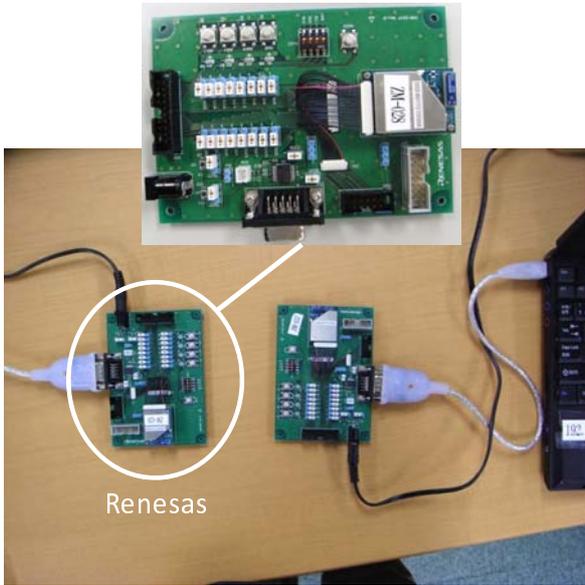


Figure 7: Snapshots of Renesas sensor nodes.

## 5 Implementation



Figure 8: Snapshots of SunSPOT sensor nodes.

We are currently implementing functionalities of the proposed coordinate integration by using actually released sensor nodes. We used a sensor node developed by Renesas [2] and SunSPOT developed by Sun Microsystems [3]. Table 2 shows the basic specifications for the sensor nodes.

Figure 7 shows the snapshots of Renesas sensor nodes. Renesas node has a main board with serial interface of RS-232C
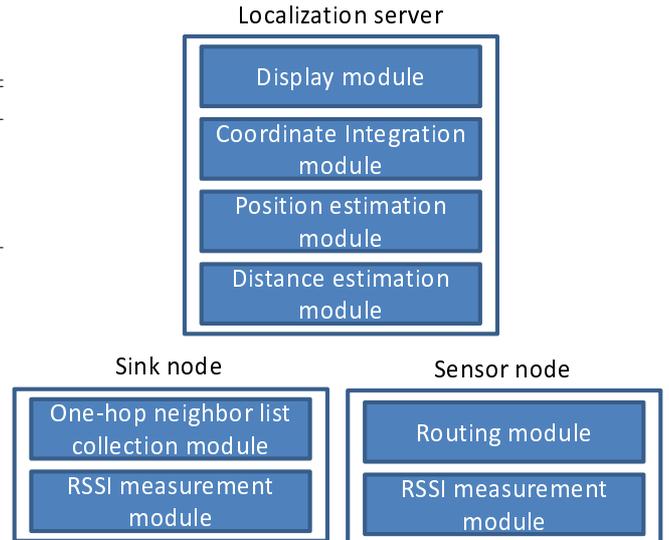


Figure 9: Required functionalities for coordinate integrations.

and equips with radio frequency (RF) modules. The RF module of Renesas node employs MC13202 developed by Freescale, and it uses IEEE 802.15.4 protocol for the PHY/MAC layers. The power is supplied by AC adapter. Renesas node provides a writable memory that a user can write the program. A user can control the microcomputer of Renesas node by writing the program. The RF module of Freescale has a functionality to store received signal strength indicator (RSSI) into registers on Renesas node. We implemented an output function of RSSI value by reading the registers of microcomputer on Renesas node.

Figure 8 shows the snapshots of SunSPOT. SunSPOT is the sensor node that can control the microcomputer by using Java program. SunSPOT has USB interface and the program can be written and debugged through the USB interface connecting to PC. SunSPOT can be operated by using battery on the board, and its power can be also supplied by using the USB cable. The RF module of SunSPOT employs CC2420 developed by Texas Instruments (TI), and it uses IEEE 802.15.4 protocol for the PHY/MAC layers. SunSPOT has **getRssi**() that is application programming interface (API) to extract the RSSI. We implemented an output function of RSSI by using the API.

Both Renesas and SunSPOT sensor nodes employ IEEE 802.15.4 protocol for PHY/MAC layer, however each protocol is only ensured to construct a sensor network within its own vender sensor nodes. One vendor sensor node cannot communicate with other vendor sensor node.

Figure 9 shows the software organization that is required to conduct the proposed integrated coordinates. Three types of nodes, i.e., localization server, sink node, and sensor node are required. We assume that a localization server execute the localization to estimate node positions. In order to conduct localization in multi-hop network, each sensor node needs to have a functionality to hold one-hop neighbor list. Each node is required to have functionalities to obtain RSSI, and relay the RSSI data to a sink node. A sink have the functionality to collect the one-hop neighbor list that is sent from sensor
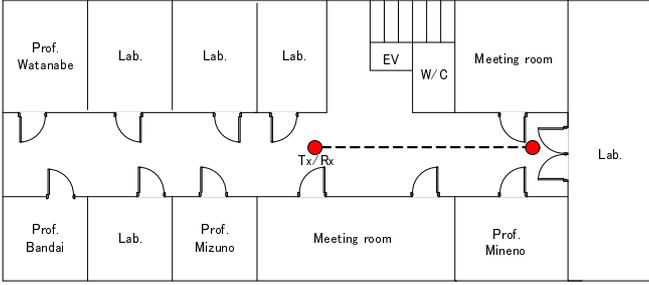
Figure 10: Floor map of RSSI experiment. Location is a corridor on the 4th floor of building in Faculty of Informatics in Hamamatsu campus of Shizuoka University, Japan.

nodes. A sink node connects to localization server, and sends the data to the server. Localization server estimates the distance from RSSI, and calculates node positions. Finally, node positions are displayed. Additionally, a functionality of data exportation of coordinates is required to implement in order to cooperate with other applications.

## 6 RSSI measurement experiment

### 6.1 Environments

In our prototype development, a sensor node uses RSSI to estimate the distance between nodes and the position. Generally, RSSI is attenuated when the node distance is longer. Then, RSSI can be used for a parameter to indicate the node distance. The advantage of using RSSI is that it can be extracted for most sensor nodes since RSSI is available if a node has radio capability. Although RSSI can be used for distance estimation, attenuation degree of RSSI depends on multiple factors such as radio frequency and measurement environments. Therefore, we conducted experiments to know how RSSI is attenuated based on node distances. Figure 10 shows a floor map that we conducted RSSI measurement experiments. The experiment location is a corridor on the 4th floor of building in Faculty of Informatics in Hamamatsu campus of Shizuoka University, Japan. RSSI was measured inside the buildings (Fig. 11(a)). The building is made by reinforced concrete. SunSPOT and Renesas nodes were closely placed parallel as shown in Fig. 11(b). We observed RSSI measurements by increasing the distance between transmitter and receiver (Fig. 11(b)). Transmission powers of both nodes were set to $-7$(dBm). We observed RSSI measurements 1000 times at each measurement points.

### 6.2 Results

Figure 12 and Fig. 13 shows results for RSSI measurement of Renesas and SunSPOT against the distance. As shown in Fig. 12, most RSSIs of Renesas node were attenuated monotonously as the node distances were longer. The maximum measurement distance to exchange the packets for RSSI measurements was 10 (m). We finalized the experiment at 10 (m) since the packet losses were too many when the node distances were over 10 (m).



(a) Corridor for RSSI measurement
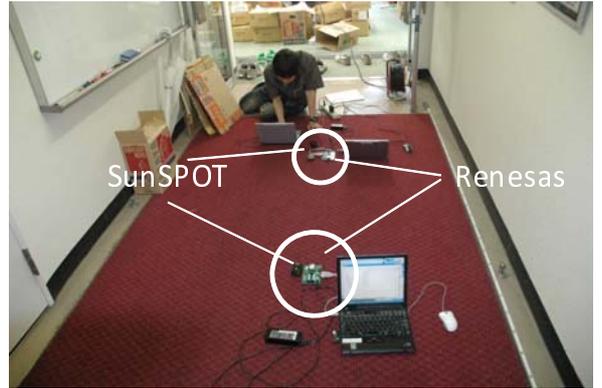


(b) Node placements for experiments

Figure 11: Snapshots of RSSI measurement environments.

It is noted that Fig. 12 and Fig. 13 appear that the number of plotted RSSI points are few. This is because RSSI values obtained from our experimental sensor platforms only read out in the form of an integer value. Much RSSI values are overlapped because of the integer format.

Figure 13 shows results for RSSI measurement of SunSPOT. RSSIs of SunSPOT were also attenuated as the node distances were longer. However, when the node distance was around 3 (m), RSSI was more attenuated than other points, which we could not observe in Renesas node even if the measurement location were same. The result suggests that attenuation of RSSI depends on each vendor of sensor node. The maximum measurement distance for SunSPOT was 7 (m).

RSSI value of Renesas at 1 (m) was $-55$ (dBm) and RSSI value of SunSPOT at 1 (m) was $-25$ (dBm) as compared in Fig. 12 and Fig. 13. The result suggests RSSI values are not compatible with vendors of sensor nodes. We need the calibration mechanism to use RSSI values for multi-vendor sensor network environments.

## 7 Summary

In this paper, we proposed the coordinate integration for heterogeneous network protocols. In the heterogeneous network environments, we revealed an issue that coordinates generated by localization techniques in one network protocol is not compatible with other coordinates generated by another network. This issue is happened when original point for coor-
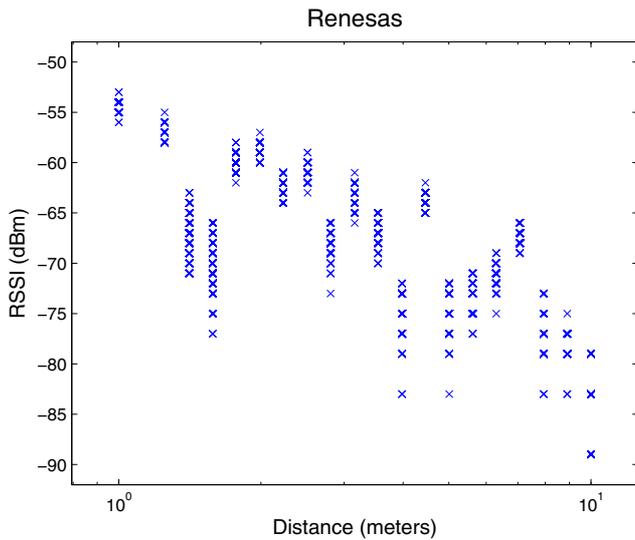
Figure 12: Experimental results of Renesas for RSSI measurements.



Figure 13: Experimental results of SunSPOT for RSSI measurements..

dinates of anchor nodes is not same as other coordinates of anchor nodes or when the coordinates are relatively determined by using anchor-free localization technique. In such environments, we described coordinate integration that each coordinates can be compatible when at least three share nodes that share the same coordinates are used. We used the simulation to verify that the operations of coordinate integration worked well. Simulation results revealed that placements for shared nodes had impacts on the positioning accuracy in proposed coordinate integrations. We also conducted RSSI measurement experiments inside the university building, and it was revealed that RSSI values were not compatible when vendors of sensor nodes were different.

We are currently implementing functionalities of proposed coordinate integration on Renesas and SunSPOT sensor nodes. In future work, we specify attenuation function of RSSI for multi-vendor sensor networks environments. We plan to evaluate the proposed coordinate integration for heterogeneous network protocol environments in the real environments.

## REFERENCES

[1] Zigbee Alliance, http://www.zigbee.org/.

[2] Renesas Technology Corp., http://japan.renesas.com/.

[3] Sun Microsystems, Inc., http://jp.sun.com/.

[4] IEEE Std 802.11–2007, "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," *IEEE Computer Society*, 2007.

[5] IEEE Std 802.15.1–2005, "Part 15.1: Wireless MAC and PHY Specifications for Wireless Personal Area Networks (WPANs)," *IEEE Computer Society*, 2005.

[6] IEEE Std 802.15.4–2006, "Part 15.4: Wireless MAC and PHY Specifications for Low-Rate WPANs," *IEEE Computer Society*, 2006.

[7] IEEE 802.15.4a task group, https://mentor.ieee.org/802.15/dcn/03/15-03-0489-03-004a-application-requirement-analysis.xls.
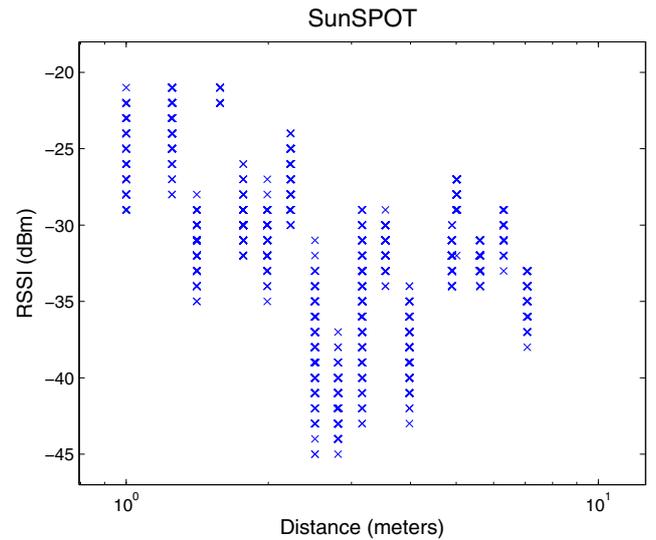
[8] M. A. Spirito, "On the Accuracy of Cellular Mobile Station Location Estimation," *IEEE Trans. on Vehicular Technology*, vol. 50, no. 3, pp. 674–685, 2001.

[9] D. Niculescu, B. Nath, "Ad Hoc Positioning System (APS)," *In Proc. of IEEE Globecom*, vol. 5, pp. 2926–2931, 2001.

[10] A. Savvides, C. Han, and M. B. Strivastava, "Dynamic Fine-grained Localization in Ad-hoc Networks of Sensors," *In Proc. of ACM Mobicom*, pp. 166–179, 2001.

[11] N. B. Priyantha, H. Balakrishnan, E. Demaine, S. Teller, "Anchor-free Distributed Localization in Sensor Networks," *MIT Technical Report 892*, Apr. 2003.

[12] Y. Shang, W. Ruml, Y. Zhang, and M. P. J. Fromherz, "Localization from Mere Connectivity," *In Proc. of ACM Mobihoc*, pp. 201–212, 2003.

[13] D. K. Goldenberg, P. Bihler, M. Cao, J. Fang, B. D. O. Anderson, A. S. Morse, and Y. R. Yang, "Localization in Sparse Networks Using Sweeps," *In Proc. of ACM Mobicom*, pp. 110–121, 2006.

[14] Jose A. Costa, Neal Patwari, and Alfred O. Hero III, "Distributed Weighted-multidimensional Scaling for Node Localization in Sensor Networks," *ACM Transactions on Sensor Networks*, vol. 2, no. 1, pp. 39–64, 2006.

[15] T. S. Rappaport, "Wireless Communications: Principles and Practice," *Prentice Hall*, 2001.

[16] J. B. Kruskal and M. Wish, "Multidimensional Scaling," *Sage Publications*, 1978.

[17] T. Masui, T. Takenaka, H. Mineno, and T. Mizuno, "Examination of Air-conditioning Contorol System Using Sensor Netowork," *In Proc.71th National Convention of Information Processing Society of Japan(IPSJ)*, pp. 3-87–3-88, Mar. 2009.

# Renewal of Pre-shared Key
# for Secure Communication of Multiple Mobile Terminals
# through Broadcast Data Distribution Systems

Hirosato Tsuji[* **], Takeshi Yoneda[**], Tadanori Mizuno[***] and Masakatsu Nishigaki[***]

[*]Graduate School of Science and Engineering, Shizuoka University, Japan
[**]Information Technology R&D Center, Mitsubishi Electric Corporation, Japan
[***]Graduate School of Science and Technology, Shizuoka University, Japan
Tsuji.Hirosato@bp.MitsubishiElectric.co.jp, Yoneda.Takeshi@ak.MitsubishiElectric.co.jp,
nisigaki@inf.shizuoka.ac.jp, mizuno@mizulab.net

***Abstract*** - To perform the secure communication of multiple mobile terminals (i.e. secure unicast communication or secure multicast communication), the encryption key should be shared among the terminals that join the communication. In such a case, if the same encryption key would be repeatly used, the key disclosure from the stolen terminal might cause the wire tapping and decryption of the encrypted communication. To minimize the risk of the disclosure, the renewal of pre-shared key must be operated. In this paper, we propose the renewal method of pre-shared key for secure communication of multiple mobile terminals. In this method, each terminal will renew its pre-shared key by one-way function (e.g. hash function) according to the instruction from the management server. We also apply the proposed method to the secure communication systems between multiple mobile terminals where the key renewal commands from the system management server to each mobile terminal is distributed through the broadcast data distribution systems.

***Keywords***: Secure Communication, Mobile Computing, Pre-shared Key Cryptography, Key Renewal, Broadcast Data Distribution Systems

## 1 INTRODUCTION

The evolution of mobile communication terminals and mobile networks enables the real-time communication using these devices. To protect against the unauthorized disclosure (i.e. wire tapping) of the communication between these terminals, the end-to-end encryption between mobiles is required. In addition, if the terminal is lost or stolen, the unauthorized use of such terminals, the decryption of encrypted communications using the stolen key from such terminals, the leakage of confidential information in such terminals should be also prevented. We've proposed the method of key/device management to realize the secure real-time communication in worldwide mobile environment [1]. In this method, the end-to-end encryption keys are frequently generated on the system management server and distributed to each terminal using one-way communication, such as the digital broadcast data distribution systems. In case of the loss or robbery of terminals, the encryption keys and the secret information will be erased and the terminal will be initialized by the remote control from the system management server. We've designed the protocol that realizes the management of encryption keys as well as the management of mobile terminals. As a result, we've confirmed that the sharing/updating encryption keys are achieved without any operation of the mobile terminal users. We've also confirmed that the lost/stolen terminal is excluded by the remote operation command of system management server and the cooperative action of the other terminals. In this paper, we propose the method renewal method of pre-shared key for secure communication of multiple mobile terminals. In this method, each terminal will renew its pre-shared key by one-way function (e.g. hash function) according to the instruction from the management server. We also apply the proposed pre-shared key renewal method to this system.

In section 2, we introduce the secure communication based on the pre-shared cryptography and threats to it. In section 3, we summarize the existing methods to protect against these threats. In section 4, we propose the method of the pre-shared key renewal. In section 5, we apply the proposed method to the secure real-time communication system. Finally, we conclude in section 6.

## 2 THREATS TO PRE-SHARED KEY CRYPTOGRAPHY

### 2.1 Secure Communication based on Pre-shared Cryptography

To perform the secure communication of multiple mobile terminals, the encryption key should be shared among the terminals that join the communication. The Figure 1 shows a method of sharing the encryption key based on the pre-shared key cryptography. In this method, the symmetric algorithm is used. Each mobile terminal has the same key shared in advance (i.e. pre-shared key) [2]. At the time of communication, each terminal derives the session key (i.e. encryption key) from its pre-shared key. Then the communication is encrypted by the derived encryption key [3].
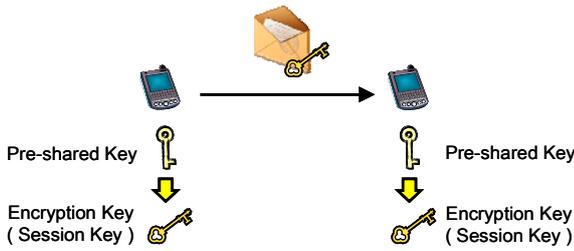
Figure 1: Secure Communication by Pre-shared Key

## 2.2 Threats from Key Disclosure

When a symmetric algorithm is used, the application of probable secure algorithms and the countermeasure against the attack on the implementation (e.g. side channel attack) are necessary [4][5][6]. They will protect against the unauthorized disclosure of the encrypted communication by the attack on the cryptographic algorithms. However, if the mobile terminal may be stolen and the encryption key can be extract from the terminal, the key disclosure might cause the following wire tapping.

**(1) Attempt to disclose past communication**
The attacker recorded the encrypted communication between terminals in advance. Then the attacker steals the one of the terminals and extracts the encryption key from it. Finally the attacker decrypts the pre-recorded communication by using the extract key.

**(2) Attempt to wiretap current/future communication**
If the terminals are used for the communication among the three or more terminals, the attacker steals one of the terminals. Then the attacker can wiretap the communication between the other terminals by using the stolen terminal. Or the attacker extracts the encryption key from the terminal in order to decrypt the communications.

The Figure 2 shows the threats to the secure communication based on the symmetric algorithm caused from the key disclosure.



Figure 2: Threat caused from Key Disclosure

Therefore the frequent renewal of the encryption key is necessary in order to decrease the risk of the disclosure of past communications. The deletion of the encryption key in the stolen terminal should be also considered to protect against the wiretapping of the current and future communications by using such terminal.

## 3   EXISTING METHODS

### 3.1  Key Renewal Methods

There are several existing methods to renew the pre-shared key for secure communication between the mobile terminals.

**(1) Pre-sharing plenty number of keys**
The plenty number of pre-shared keys should be generated in advance. Then these keys had been pre-installed to each mobile terminal. The renewals of pre-shared key will be performed at the sufficient cycles.

**(2) Renewal Key Establishment without Servers** [2]
When the renewal of pre-shared key is required, the mobile terminals communicate each other to establish the new pre-shared keys. In general, the asymmetric algorithm is used such as RSA key transfer algorithm or Diffie-Hellman key agreement algorithm. The combination with the digital signature algorithm such as DSA signature algorithm or RSA signature algorithm is must be also performed.

**(3) Renewal Key Establishment with Servers** [7]
When the renewal of pre-shared key is required, the mobile terminals communicate to the trusted server to establish the new pre-shared keys. In general, the server acts as the key distribution center (KDC) and generates the new pre-shared keys and distributes them to each mobile terminals. The well-known implementation of KDC is the Kerberos server.

### 3.2  Terminal Management Methods

There are several existing methods to protect against the unauthorized use of the stolen mobile terminals.

**(1) User Authentication**
The mobile terminal can be protected from the unauthorized use by the user authentication functions. The authentication information may be PIN, password and the biometrics authentication information (e.g. fingerprint).

**(2) Tamper Resistant Terminal**
The extraction of keys from the mobile terminal can be protected, if the mobile terminal has the tamper resistant function. The keys will be automatically erased when the stolen mobile terminal would be illegally opened.

**(3) Remote Management from Server** [8]
The mobile terminals are managed by the management server. When the mobile terminal would be stolen, the remote operation command will be sent from the management server to the stolen terminal. Then the stolen terminal will be locked or initialized. In the latter case, the keys in the stolen terminal are erased.

# 4 PROPOSED METHOD

## 4.1 Design Policy

We propose the method of the pre-shared encryption key management for the secure communication of multiple mobile terminals. The design policies of our proposal are the followings.

### (1) Total Management by Administrator

The management (i.e. renewal and deletion) of the pre-shared key should be performed according to the order of the system administrator. The users are not allowed to manage any pre-shared keys without the permission of the administrator.

### (2) No distribution of Renewal Key

At the time of key renewal, the updated pre-shared key itself will not be distributed nor transferred via network.

### (3) Deletion of Key in the stolen Terminal

In case that the loss or robbery of mobile terminal might be happen, the keys in such terminal will be deleted to avoid the unauthorized disclosure of the secure communication.

## 4.2 System Architecture

The Figure 3 shows the basic system architecture of the proposed method. The components of the system are defined as the followings.

- **User Terminal**
  A terminal that perform the secure communication with other terminals using the pre-shared key.
- **Management Server**
  A server to manage the whole pre-shared key among user terminals.
- **Pre-shared Key**
  A key shared in advance among the user terminals. The pre-shared key consists of the following elements.

| Element | Meanings |
|---|---|
| Key ID | Specify the identifier of Key |
| Key | Encryption Key |
| Generation Number | Specify the current generation of key renewal |
| Validity Period | Specify the next key renewal date/time (optional element) |

- **Key Generation Renewal Command**
  A command sent from the management server to each user terminal in order to dictate the renewal of pre-shared key. The authentication value (e.g. the digital signature or the Message Authentication Code) is added to the command for the verification.
- **Pre-shared Key Renewal Mechanism**
  A kind of cryptographic mechanism implemented on both the management server and user terminals. The input is the current pre-shared key and the output is the next generation of pre-shared key.

- **Key Deletion Command**
  A command sent from the management server to the specific user terminal in order to delete the keys in the stolen terminal. The authentication value is added to the command for the verification.
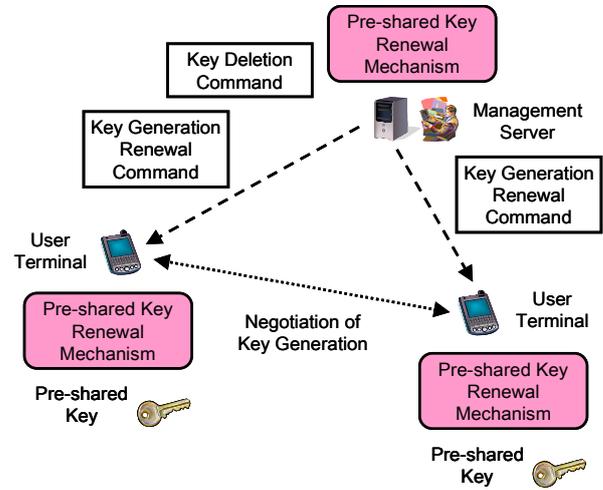


Figure 3: System Architecture of Proposed Method

## 4.3 Renewal of Pre-shared Key

The pre-shared key is renewed according to one of the following instructions.

### (1) Renewal by Key Renewal Command from Server

The management server creates the key generation renewal command and issues it to the user terminals. The user terminal that receive the command will verify it and renew own pre-shared key.

### (2) Automatic Renewal by Key Validity Period

The management server creates and issues the key generation renewal command with the key validity period. The user terminal will automatically renew its own pre-shared key when the key validity period would be expired.

### (3) Synchronized Renewal in case of Generation Gap

The key generation renewal command from the management server may be lost. If it happens, the pre-shared key of the user terminal might not be renewed. When the key generation gap between the user terminals would be detected at the beginning of the secure communication, the user terminal will automatically renew the old pre-shared key for synchronization.

## 4.4 Renewal Mechanism Example

The pre-shared key renewal mechanism in both the management server and the user terminals can be implemented by using the probable secure one way functions (e.g. hush functions). The figure 4 shows the example of the implementation, where the current pre-shared key has been inputted and the next generation of pre-shared key will be outputted.

Figure 4: Pre-shared Key Renewal Mechanism

# 5 APPLICATIONS

## 5.1 Encryption Key Management through Broadcast Data Distribution Systems

We've proposed the secure real-time communication system where the pre-shared key for the end-to-end encryption between mobile terminals are generated by the system management server and are distributed through the broadcast data distribution systems in [1] (Figure 5). In this paper, we apply the proposed pre-shared key renewal method to this system.
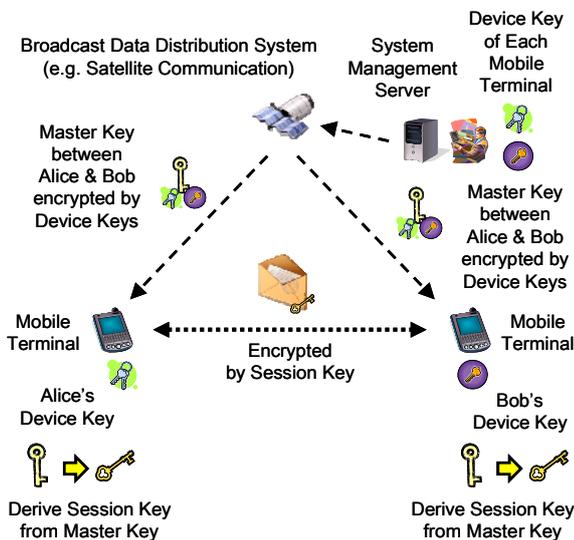


Figure 5: End-to-End Encryption Key Management through Broadcast Data Distribution Systems

This secure real-time communication system consists of the following entities/elements.

- **Mobile Terminal**
  A terminal that performs the secure real-time communication of voices and movies. To protect against the unauthorized disclosure, the end-to-end encryption is operated.
- **System Management Server**
  A trusted server that generates and distributes the pre-shared keys used for the secure communication between mobile terminals. It also orders the renewal of pre-shared keys among the mobile terminals and the deletion of keys in the stolen mobile terminal.
- **Broadcast Data Distribution System**
  A network used for the one-way communication from the system management server to each mobile terminal, such as the broadcast data distribution service using the satellite communication.

The following three types of keys are used in this system.

- **Device Key**
  A pre-shared key between each mobile terminal and the system management server. A device key is used to encrypt the commands that are sent from the system management server to each terminal. It is also used to archive integrity of the commands.
- **Master Key**
  A pre-shared key between mobile terminals that perform secure communication. An initial master key is generated in the system management server and distributed to each mobile terminal via a key distribution command.
- **Session Key**
  An encryption key between mobile terminals that perform secure communication. A session key is derived from the master key shared in advance.

## 5.2 Pre-sharing Device Key

The device keys must be shared between the system management server and each mobile terminal before the operation of the secure communication system would be started. The system management server generates the device keys for each mobile terminal. Then each device key was pre-distributed to the corresponding terminals securely. The figure 6 shows the pre-sharing device key between the system management server and the mobile terminals.
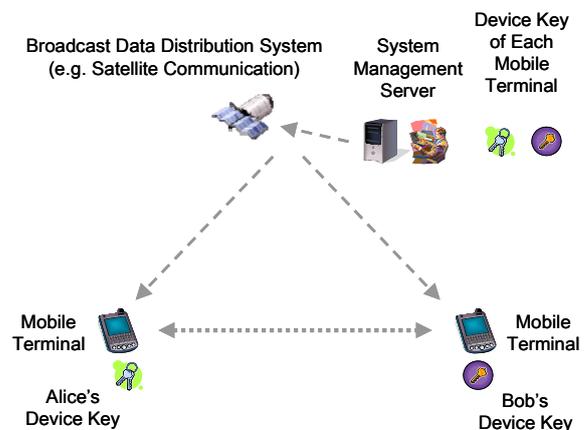


Figure 6: Pre-sharing Device Key

## 5.3 Distribution of Master Key

The system management server generates the master keys which are used for the secure communication between mobile terminals. Then the master keys are encrypted by the pre-shared device keys where the only legal terminal can decrypt the encrypted master key. Then system management server distributes the encrypted master keys via the broadcast data distribution systems to the whole mobile terminals. For example, the system management server generates the master key between Alice and Bob. The server generates the key-encryption key and encrypts the master key by the key-encryption key. The both encryptions of the key-encryption key by Alice's device key and Bob's device key are added to the master key. That is, the master key can

be only decrypted by either Alice's device key or Bob's device key. The master key for Alice and Bob is distributed through the satellite broadcast data distribution system. All mobile terminals including Carol's receive the encrypted master key, but only Alice's and Bobs' can be decrypt it. The figure 7 shows the generation and distribution of master keys.
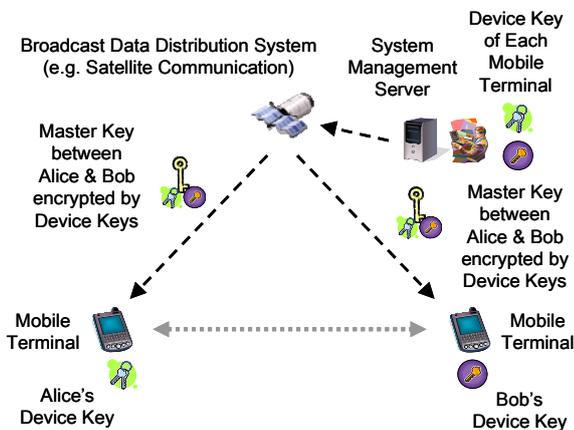


Figure 7: Generation and Distribution of Master Key

## 5.4 Generation of Session Key

At the time of secure communication, each mobile terminal derives the session key from the pre-shared master key. Then their communication is encrypted by the session key. The figure 8 shows the generation of session key and secure communication with it. The session key should be re-derived from the same master key when it would be used to encrypt/decrypt the specific number of times.
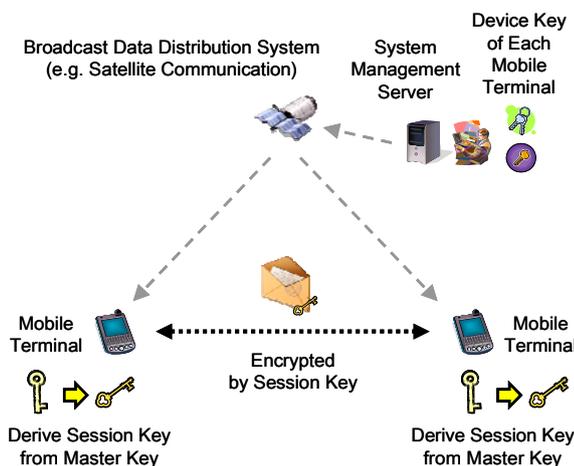


Figure 8: Generation of Session Key

## 5.5 Renewal of Device Key

The device keys are used for the one-way secure communication from the system management server to each mobile terminal. Therefore the keys should be renewed at the sufficient intervals. The renewal of device keys are operated according to the instructions described in section 4.3. The figure 9 shows the renewal of device keys by

sending the device key renewal command from the system management server to each mobile terminal.
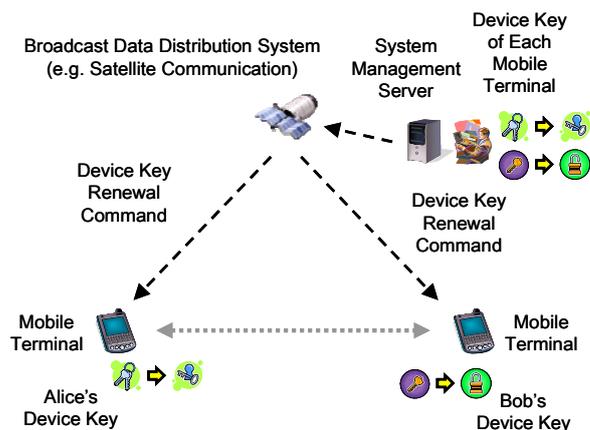


Figure 9: Renewal of Device Key

The mobile terminals will also renew the device keys when the validity of the current device key would be expired or when the generation gap of the device keys would be detected.

## 5.6 Renewal of Master Key

The master keys are used to derive the session keys (encryption keys) between mobile terminals. The master key should be renewed at the sufficient intervals. In some cases, the master keys must be renewed at every time when the new secure session would be established between the same pair of mobile terminals. Thus the renewal of master keys are operated according to the instructions described in section 4.3. The figure 10 shows the renewal of master keys by sending the master key renewal command from the system management server to each mobile terminal.
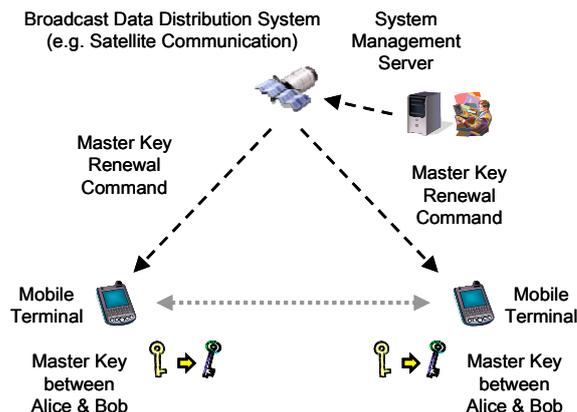


Figure 10: Renewal of Master Key (1)

The mobile terminals will also renew the master keys when the validity of the current master key would be expired or when the generation gap of the master keys would be detected. The figure 11 shows another renewal of master keys by the detection of generation gap between the mobile terminals.
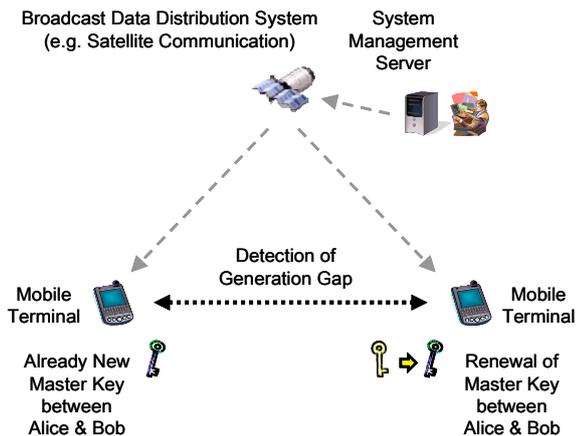
Figure 11: Renewal of Master Key (2)

The renewal of master keys may be achieved by the re-distribution of new master keys from the system management server. If the distribution of renewal key (see 3.1) may be acceptable, the new master keys might be distributed by the same procedure as mentioned in section 5.3.

## 5.7 Deletion of Keys

If the mobile terminal might be stolen, the user would report that to the system administrator. The administrator will operate the system management server in order to create the key deletion command and send it to the stolen terminal. The terminal that receives the command will delete the all cryptographic keys inside. In some cases, the initialization of the mobile terminal (the deletion of everything inside the terminal) may be performed by sending the terminal initialization command instead. The figure 12 shows the deletion of keys in the stolen terminal.
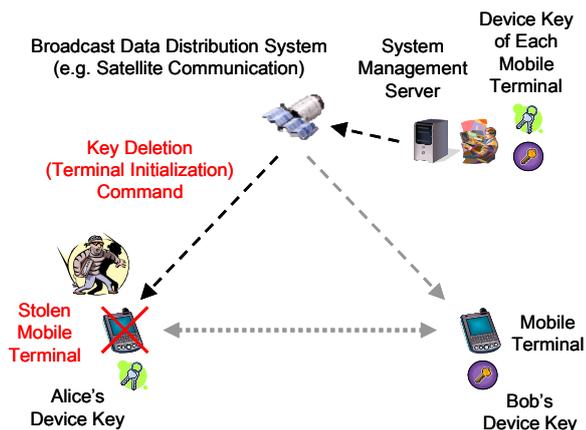


Figure 12: Deletion of Keys in stolen Terminal

## 6 CONCLUSION

In this paper, we propose the renewal method of the pre-shared key for the secure communication of multiple mobile terminals. We also apply the proposed method to the secure real-time communication systems where the management operations from the system management server to the mobile terminals are distributed through the broadcast data distribution systems. The prototype of this system is being implemented, but not completed yet [9]. We will extend our implementation to support the proposed method and evaluate its effectiveness.

## REFERENCES

[1] H.Tsuji, T.Yoneda, T.Mizuno and M.Nishigaki, Realization of Secure Real-time Communication for Worldwide Mobile Environment by Frequent Key Renewal Method through Broadcast Data Distribution Systems, IPSJ Journal Vol.50 No.9 (2009).

[2] RFC 3830, MIKEY: Multimedia Internet KEYing (2004).

[3] RFC 3711, The Secure Real-time Transport Protocol (SRTP) (2004).

[4] P.Kocher, Timing Attacks on Implementation of Diffie-Hellman, RSA, DSS and Other Systems, CRYPTO'96 (1996).

[5] P.Kocher, J.Jaffe, B.Jun, Differential Power Analysis, CRYPTO'99 (1999).

[6] Y.Tsunoo, E.Tsujihara, K.Minematsu, H.Miyauchi, Cryptoanalysis of Block Ciphers Implemented on Computers with Cache, ISITA2002 (2002).

[7] B.C.Neuman, T.Ts'o: Kerberos: An Authentication Service for Computer Networks, IEEE Communications Magazine, Vol.32, No.9 (1994)

[8] Open Mobile Alliance: OMA Device Management Protocol, Approved Version 1.2.1 - 17 Jun 2008 (2008).

[9] H.Tsuji and T.Yoneda, Secure Mobile Phone System, Mitsubishi Electric Technical Report, Vol.82 No.5 (2008)

[10] H.Tsuji and T.Yoneda, Renewal of Pre-shared Key among Multiple Mobile Terminals, The 2009 Symposium on Cryptography and Information Security (SCIS 2009), 3D4-3 (2009).

# An Algorithm for Assigning Real-time Tasks with Timing Constraints into Dynamically Reconfigurable Processor

Tomoya Kitani[†], Keisuke Nishi[‡], and Teruo Higashino[‡]
[†]Shizuoka University, Japan    [‡]Osaka Univeristy, Japan

*Abstract* - In this paper, we focus on the problem of implementing a periodic concurrent system with timing constraints into multi-context dynamically reconfigurable processors (DRP). A concurrent system has multiple tasks that can be executed in parallel. Moreover, some tasks in a specific set of processes might be required to synchronize each other. We propose a method for assigning tasks into a multi-context DRP such that timing constraints of the system are satisfied and the size of the program area required on each context for implementing the given system is minimized. However this problem is a combinatorial problem. Thus we propose a heuristic algorithm for solving the problem efficiently. Experimental results show that the proposed method can derive a quasi-optimal assignment in a short time.

*Keywords*: real-time system, dynamically reconfigurable processor, task assignment, heuristic algorithm

## 1 Introduction

Recently it has been required to manufacture a wide variety of embedded systems in small quantities in a short period of time. For the purpose, many embedded systems have been implemented to reconfigurable devices. With the advent of modern field programmable devices, many systems are implemented using small sized reconfigurable devices such as Dynamically Reconfigurable Processors (DRPs). These devices have many logical areas and can utilize these areas efficiently. A system can be implemented into a DRP by decomposing and allocating its multiple contexts into its logical areas.

Many systems are generally real-time systems with timing constraints. A proper decomposition or implementation of a real-time system into a multi-context DRP requires that the decomposed contexts satisfy the timing constraints of the given system. This is usually a hard task since the execution of decomposed contexts involves context switching and proper scheduling by considering synchronization constraints of the system. Moreover, a proper decomposition can take into account reducing the size of the program area required on each context for implementing the given system.

In this paper, a behavior (or a specification) of a system is assumed to be given as a set of periodic concurrent processes with timing constraints [1], [2]. Periodic systems are used in many real-time application areas such as communication systems, multimedia systems, routers and so on [3]–[5]. We assume that all processes have the same time period $N_t$. A process consists of a finite length of sequences of tasks, which can be represented in the form of a tree structure. A task corresponds to a functional module of the system. We assume that each task has several attributes such as its starting time,

execution time, size (representing the program area necessary for implementing the task), timing constraint, temporal ordering relation with other tasks in the same process, and synchronizing condition with specific processes in the system.

We focus on a problem of decomposing a given periodic concurrent system and assigning its tasks into a multi-context DRP. The multi-context DRP can change its contexts dynamically in order to switch executable sub-modules. All the tasks on the same context can be executed in parallel. However, tasks assigned to different contexts cannot be executed simultaneously. Thus, we need assign those tasks carefully into a multi-context DRP so that all the tasks do not violate their timing constraints. We present a method for decomposing a given concurrent system and assigning its tasks into a multi-context DRP in such a way that the timing constraints of the system hold and that the maximum size of the programmable logic areas used on the target contexts is minimized. We propose a heuristic algorithm to solve this problem since it is a combinatorial problem. The proposed algorithm consists of three steps. First, given a system, the algorithm derives all tasks which satisfy the timing constraints of the system. Next, it constructs a task dependence graph about tasks' execution time range satisfying their timing constraints. Then, it assigns the tasks into contexts of a DRP according to the graph.

In the experimental results, we have confirmed that the proposed heuristic algorithm derives quasi-optical assignment results for large size examples in short time.

## 2 Related Work

A lot of research work has been carried out on partitioning and scheduling systems into reconfigurable systems [6]–[13]. The work in [6] targets a general multi-context reconfigurable architecture and focuses on context scheduling considering the overheads of context switching.

The work in [7] deals with temporal partitioning and scheduling of data flow graphs into reconfigurable computers. In particular, the work considers minimizing the additional number of configurable logic blocks of an FPGA.

In [8] a method for the synthesis and temporal partitioning of reconfigurable systems is provided. The method considers minimizing the number of ports of a DRP internal memory. In [9], a 0-1 non-linear programming model is given for temporal partitioning and high-level synthesis for implementing a system of reconfigurable processors. The model considers minimizing the size of data communication among partitioned tasks. A data scheduler for multi-context reconfigurable architectures is provided in [10]. The scheduler minimizes the size of data communication between DRP's external memory and its Programmable Logic Area (PLA).

In [11] and [12] partitioning methods are presented for reconfigurable systems considering parallelization of tasks, accurate reconfiguration overhead, and minimization of the total computation time of the system. In [13] a method for task scheduling with configuration prefetching during reconfiguration overhead is given. An ILP model for minimizing the total computation time of the system is provided. The model takes into account reconfiguration overhead and a given limit on the size of the programmable logic area.

Similar to our work, the above described work targets reconfigurable architectures. However, the system models considered in the above work consider order relations among tasks for tasks with no timing constraints. The work given in [5] uses a similar task model as the one presented in this paper, i.e. concurrent, periodic, and with time constraints. However, unlike our work, the work given in [5] does not consider reconfigurable architectures. In this paper, we use concurrent periodic task model with timing constraints over multi-context reconfigurable processors.

# 3 Target concurrent system and DRP

## 3.1 Real-time system model

### 3.1.1 Definition of system

In this paper, a concurrent periodic real-time system is described as a 4-tuple $System = (Processes, Tasks, clock, N_t)$. $Processes$ denotes a set of processes. $Tasks$ denotes a set of tasks used in the system. $clock$ is an integer counter that counts the elapsed time from the initial state of the system, and $N_t$ is an integer that denotes the time period of the system.

A process in $Processes$ consists of a finite length of sequences of tasks in $Tasks$ that can be represented in the form of a tree structure. The root node of a tree structure represents the initial state. A sequence of consecutive tasks starting from the initial state (root node) and ending at a leaf node is called a *path*. All processes of the system have the same time period $N_t$. Each process executes tasks in one of its paths until it reaches the end of the path. If the time period $N_t$ has passed before reaching the end of the path, $clock$ is reset to $0$ and the process returns to its initial state. Note that we assume any process does not have loops. A simple example of a concurrent system System1 is shown in Fig. 1. System1 consists of 2 processes: Process1 and Process2. Process1 consists of two paths {P11 and P12}, and Process2 consists of two paths {P21 and P22}.

In our model, we assume some tasks can synchronize among a specific set of processes. The synchronization relation as well as the behavior of each process is formally specified as follows using the operators in LOTOS language [14].

$$S ::= (S\,|[task\_list]|\,S)\ \ |\ \ (S\,|||\,S)\ \ |\ \ P$$
$$P ::=\ \ task;\,P\ \ |\ \ (P[]P)\ \ ||\ \ task$$

Here, $S$ denotes a concurrent system, and $P$ denotes a process. In $S$, $|[task\_list]|$ denotes the synchronous parallel operator where $task\_list$ is a list of tasks to be synchronized
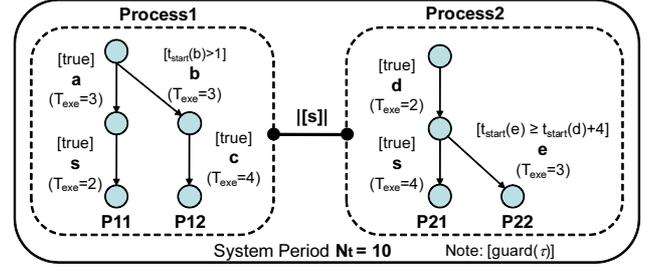


Figure 1: An example of a system, System1

between its operator's both sides of processes. If the same task $s$ is specified in two processes $P1$ and $P2$, and if those processes are combined with the synchronous parallel operator $P1|[s]|P2$, the task $s$ in $P1$ and $P2$ must be executed simultaneously. The operator $|||$ is the asynchronous parallel operator, and it denotes that its operator's both sides of processes can run in parallel without any synchronization. In $P$, the operator $[]$ is the choice operator, and $(P1[]P2)$ denotes that either $P1$ or $P2$ is executed. The choice operator $[]$ corresponds to a branch in a given process's tree structure. $task; P$ denotes the sequential composition of tasks.

Each task $\tau$ has a variable $t_{start}(\tau)$ and four attributes $(T_{exe}(\tau), \tau_{prev}(\tau), guard(\tau), size(\tau))$. Although each task may appear in many paths of a given process if it can appear only once in any path of the process, here we assume that each task can appear only once in each process. The integer variable $t_{start}(\tau)$ denotes the starting time of task $\tau$. $T_{exe}(\tau)$ is an integer value denoting the duration (or execution time) of task $\tau$. Here, we assume the execution of a task is not interrupted by other tasks. $\tau_{prev}(\tau)$ denotes the previous task of $\tau$. A $guard(\tau)$ denotes timing constraints that $\tau$ has to satisfy, and it is represented as a logical product of timing constraints, each of which is represented as a linear inequality over variables $\{t_{start}(\tau')|\tau'$ is an ancestor task of $\tau\}$. $size(\tau)$ denotes the size of $\tau$, and it corresponds to the size of the program area necessary for implementing $\tau$.

In (Process1|[s]|Process2) in Fig. 1, task s in Process1 and Process2 must be executed simultaneously. When the synchronous task s is executed simultaneously, both the guards in the two processes must hold. Given a system of $n$ processes, $pr_1, \cdots, pr_n$, each of which has a set of finite lengths' paths $Paths^{pr_1}, \cdots, Paths^{pr_n}$, a *path-combination* is a list of $n$ paths $\{< p_1, \cdots, p_n > \ |\ p_1 \in Paths^{pr_1}, \cdots, p_n \in Paths^{pr_n}\}$. All (syntactical) path combinations in Fig. 1 are m1:<P11,P21>, m2:<P11,P22>, m3:<P12,P21>, and m4:<P12,P22>.

However, the path combinations m2 and m3 cannot be executed until their leaf nodes since the synchronous task s cannot be synchronized between two processes. The path combinations m1 and m4 are syntactically possible path combinations. We describe about behavior of a system as follows.

### 3.1.2 Behavior of system

A task $\tau$ is *valid* (*executable*) in a path-combination $m$, denoted $valid_{task}(\tau, m)$, if and only if it satisfies the following

three constraints:

The previous task $\tau_{prev}(\tau)$ has finished its execution.

$guard(\tau)$ is true.

$\tau$ ends its execution before the time period $N_t$.

Those constraints can be represented using the following linear constraints:

$$\begin{cases} guard(\tau) \wedge (t_{start}(\tau) + T_{exe}(\tau) \leq N_t) \\ \quad (\tau \text{ is the initial task of a path}). \\ (t_{start}(\tau') + T_{exe}(\tau') \leq t_{start}(\tau)) \wedge guard(\tau) \\ \quad \wedge (t_{start}(\tau) + T_{exe}(\tau) \leq N_t) \\ \quad (\text{otherwise}) \ (\tau' \text{ denotes } \tau_{prev}(\tau)). \end{cases}$$

A *valid path* of a process consists of a sequence of only valid tasks.

If a path combination $m$ which consists of valid paths satisfies the following conditions, $m$ is called a *valid path combination*:

$m$ contains all synchronized tasks $Task_{sync}(\tau)$ of each task $\tau$ in $m$.

Each task $\tau$ which should be synchronized with tasks can start at the time when the synchronized tasks start.

The latter constraint can be represented using the following linear constraint:

$$\forall \tau' \in Task_{sync}(\tau) \, (t_{start}(\tau) = t_{start}(\tau'))$$

For this reason, a path combination which consists of only valid paths could not be valid.

A system is execute a valid path combination each period. If the system executes tasks in any invalid path combination, the system cannot reach their leaf nodes before the time period $N_t$ has passed. This corresponds to either case of (i) a deadlock occurs, or (ii) a given task cannot be executed before the time period $N_t$ has passed. Here, we focus on the implementation of all tasks in all valid path combinations of a given concurrent system.

## 3.2 Dynamically Reconfigurable Processors

Recent DRP consists of multiple contexts with several tiny processors called Processing Elements (PEs). The contexts of a DRP represent logical configurations of a PLA, and the DRP can switch its contexts dynamically with small overheads. Recently, many DRP's have been developed by companies and research institutes [15]–[19]. NEC Elect. Corp. has developed a multi-context DRP with the same program area size [15]. IPFlex Inc. has developed DAPDANA-2 [16], Singh et al. at U.C. Irvine have developed MorphoSys [17], Goldstein et al. in Carnegie Mellon Univ. have developed PipeRench [18], and Mei et al. in IMEC have developed ADRES [19]. DRPs can implement the systems at their programmable logic areas efficiently. Additionally, DRPs can hold many configurations, which are context assignments of a given system, without changing their logical area size. For example,

the system implemented into a DRP can handle computation load flexibly by utilizing several configurations that mount the same function but with different performance and power consumption.

In this paper, we adopt a multi-context DRP architecture with the same program area size and a constant switching overhead, which is close to the multi-context DRP model of NEC Elect. Corp. in [15]. We define the target multi-context DRP as $DRP = (N_c, max_{cont}, T_{switch})$, where $N_c$ denotes the number of contexts of the DRP, and $max_{cont}$ and $T_{switch}$ denote the size limit of the contexts and the context switching overhead, respectively.

When we implement a concurrent system defined above on the target multi-context DRP, it might occur a case that we cannot execute all valid path combinations of a given concurrent system due to the shortage of contexts, the context switching overhead and/or the size limit of contexts. For a given concurrent system $Sys$, if an implementation $Imp$ on a multi-context DRP can execute all valid path combinations in $Sys$, then we say that $Imp$ is a correct implementation of $Sys$. Hereafter, we discuss about a method to derive a correct implementation $Imp$ on a target multi-context DRP from a given concurrent system $Sys$.



(a) Architecture of DRP     (b) Concept of Context Switching
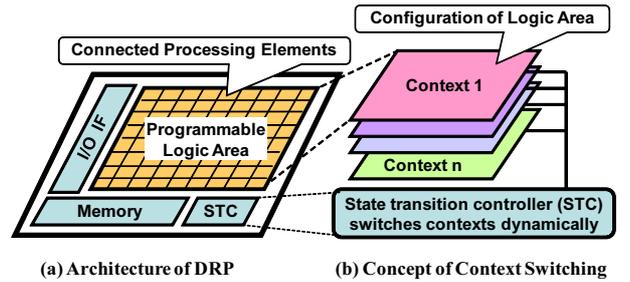
Figure 2: An architecture of DRP

## 4 System Decomposition into DRP

A system is decomposed into contexts of a DRP in order to implement a system on DRP. The decomposed system should also satisfy the timing constraints of the system.

In this section, we propose a heuristic algorithm to assign tasks in a system into contexts of a DRP so that all the tasks execute satisfying their timing constraints.

### 4.1 Problem Definition

Inputs:

A system $System = (Task, clock, N_t)$, and a DRP $DRP = (N_c, T_{switch})$.

Output:

$N_c$ sets of tasks each of which is assigned into a context.

Constraints:

A set of valid path-combinations of the system is invariant whenever the system is assigned into the DRP and all the valid path-combinations are executable.

Objective:
To minimize the size of contexts of the DRP.

## 4.2 Proposed Heuristic Algorithm

The proposed algorithm consists of three parts: (1) to derive all tasks in valid path combinations, (2) to construct a task dependence graph about tasks' valid time range, and (3) to assign tasks into contexts according to the graph. For simplification of this problem, we assume that the context switching overhead $T_{switch}$ is zero. In our future work, we will consider about the overhead.

### 4.2.1 Valid Time Range of Each Task

First, we derive all tasks in valid path combinations of a given system.

Given a task $\tau$ in a path $p$, a *valid time range* of $\tau$ in $p$, denoted $vrange(\tau, p)$, is defined as the range between the earliest startable time $t_{est}(\tau, p)$ of $\tau$ in $p$, and the latest startable time $t_{lst}(\tau, p)$ of $\tau$ in $p$. The earliest starting time $t_{est}(\tau, p)$ is calculated as the minimal value that makes $valid_{task}(\tau, p)$ true in $p$. Moreover, if $\tau$ is the last task of $p$, then $t_{lst}(\tau, p)$ is calculated as the maximal value that makes $valid_{task}(\tau, p)$ true in $p$. Otherwise, given the task $\tau'$ of $p$ and its previous (parent) task $\tau$ of $\tau'$, $t_{lst}(\tau, p)$ is calculated as the maximal value of (a) the maximal value of $t_{start}(\tau)$ that makes $valid_{task}(\tau, p)$ true in $p$ and (b) $(t_{lst}(\tau') - T_{exe}(\tau))$. An example of the valid time range of tasks in a path is shown in Fig. 3.



(a) A valid path **p1**       (b) A valid time range
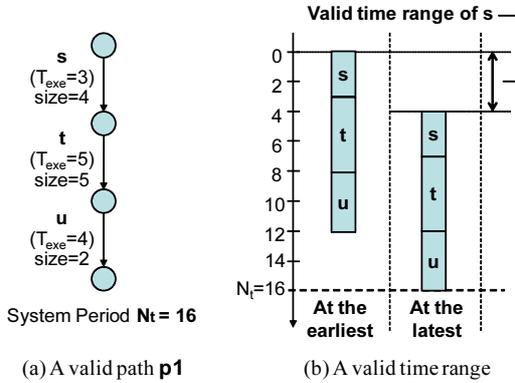
Figure 3: Valid time range

Given a path $p$ and a task $\tau$ in $p$, the earliest startable time $t_{est}(\tau, p)$ and the latest startable time $t_{lst}(\tau, p)$ are calculated from $guard(\tau)$ as follows. Here, $\tau$ and $\tau'$ are tasks in a path $p$, $k$ denotes a coefficient value, and $D$ denotes a set of terms of the other tasks' variables and constant terms. If an inequality in $guard(\tau)$ is expressed as $t_{start}(\tau) < k \cdot t_{start}(\tau') + D$ then

$$t_{lst}(\tau, p) \leftarrow \min\left(t_{lst}(\tau, p), k \cdot t_{lst}(\tau', p) + D\right) \quad \text{and} \tag{1a}$$

$$t_{est}(\tau', p) \leftarrow \max\left(t_{est}(\tau', p), \frac{t_{lst}(\tau, p) - D}{k}\right), \tag{1b}$$

$t_{start}(\tau, p) < D - k \cdot t_{start}(\tau', p)$ then

$$t_{lst}(\tau, p) \leftarrow \min\left(t_{lst}(\tau, p), D - k \cdot t_{est}(\tau', p)\right) \quad \text{and} \tag{2a}$$

$$t_{lst}(\tau', p) \leftarrow \min\left(t_{lst}(\tau', p), \frac{D - t_{est}(\tau, p)}{k}\right), \tag{2b}$$

$t_{start}(\tau) > k \cdot t_{start}(\tau') + D$ then

$$t_{est}(\tau, p) \leftarrow \max\left(t_{est}(\tau, p), k \cdot t_{est}(\tau', p) + D\right) \quad \text{and} \tag{3a}$$

$$t_{lst}(\tau', p) \leftarrow \min\left(t_{lst}(\tau', p), \frac{t_{lst}(\tau, p) - D}{k}\right), \tag{3b}$$

or $t_{start}(\tau) > D - k \cdot t_{start}(\tau')$ then

$$t_{est}(\tau, p) \leftarrow \max\left(t_{est}(\tau, p), D - k \cdot t_{lst}(\tau', p)\right) \quad \text{and} \tag{4a}$$

$$t_{est}(\tau', p) \leftarrow \max\left(t_{est}(\tau', p), \frac{D - t_{lst}(\tau, p)}{k}\right). \tag{4b}$$

A task $\tau$ in a path $p$ is valid if $t_{est}(\tau, p)$ and $t_{lst}(\tau, p)$ are defined (i.e., they have values) and $(t_{est}(\tau, p) \leq t_{lst}(\tau, p))$. Otherwise, task $\tau$ in path $p$ is invalid, and $p$ is also invalid.

Next, we consider about synchronizations among tasks. A task $\tau$ and a set of synchronized tasks $Task_{sync}(\tau)$ with $\tau$ should start their execution simultaneously. Thus, given $\tau$ and $Task_{sync}(\tau)$, $t_est(\tau, p)$ and $t_lst(\tau, p)$ are recalculated as follows:

$$t_{est}(\tau', p') \leftarrow \max\left(t_{est}(\tau, p), \forall \tau''(t_{est}(\tau'', p''))\right), \quad \text{and} \tag{5a}$$

$$t_{lst}(\tau', p') \leftarrow \min\left(t_{lst}(\tau, p), \forall \tau''(t_{lst}(\tau'', p''))\right). \tag{5b}$$

Here, $\tau', \tau'' \in \{\tau\} \cup Task_{sync}(\tau)$, and $p'$ and $p''$ denote paths which contain $\tau'$ and $\tau''$, respectively. $t_{est}(\tau, p)$ and $t_{lst}(\tau, p)$ of all tasks in a path are updated as recalculated by Eqs. (1a)–(4b) whenever at least one of them in the path is recalculated. A valid path consists of valid tasks after the recalculation about synchronization.

Finally, we derive a *valid process*. Given a process $pr$, a task $\tau$ in $pr$ and all the valid paths $p_1, \cdots p_n$ which contain $\tau$ in $pr$, $t_est(\tau, p_i)$ and $t_lst(\tau, p_i)$ $(i \in \{1, \cdots, n\})$ are recalculated as follows:

$$t_{est}(\tau, p_i) \leftarrow \max_{j \in \{1, \cdots, n\}}\left(t_{est}(\tau, p_j)\right), \quad \text{and} \tag{6a}$$

$$t_{lst}(\tau, p_i) \leftarrow \min_{j \in \{1, \cdots, n\}}\left(t_{lst}(\tau, p_j)\right). \tag{6b}$$

Here, $j \in \{1, \cdots, n\}$. $t_{est}(\tau, p)$ and $t_{lst}(\tau, p)$ of all tasks in a path are updated too whenever at least one of them in the path is recalculated. A valid process consists of valid paths after the recalculation about .

A *valid system* consists of valid processes. In each period a valid system executes one of path combinations which consist of paths in each valid process one by one.

#### 4.2.2 Construction of Task Dependency Graph about Valid Time Range

In a valid system, a task starts its execution within its valid time range. $st(\tau)$ denotes the time when an implemented task $\tau$ into a DRP starts its execution. We call $st(\tau)$ a scheduled time of $\tau$. We can decide $st(\tau)$ within the valid time range of $\tau$. We call this decision of $st(\tau)$ a *scheduling*. A task $\tau$ is running on a DRP between the time $st(\tau)$ and $st(\tau) + T_{exe}(\tau)$. We call this time range a *running time range*.

Given two tasks $\tau_1$ and $\tau_2$, they should be assigned into a same context of a DRP if the running time range of $\tau_1$ and that of $\tau_2$ overlap. On the other hand, if the running time ranges do not overlap, $\tau_1$ can be assigned into different contexts from $\tau_2$ so that the maximum size of contexts is reduced.

Given a task $\tau$ in a path $p$, the *earliest end time* and the *latest end time* of $\tau$ in $p$ are calculated as $t_{est}(\tau, p) + T_{exe}(\tau)$ and $t_{lst}(\tau, p) + T_{exe}(\tau)$, respectively, as shown in Fig. 4. The running time range of $\tau$ depends on the scheduling of $st(\tau)$. However, the time range when $\tau$ is surely running can exist regardless of the scheduling of $\tau$. We call such time range a *surely running range* as shown in the figure. If $t_{lst}(\tau) \leq t_{est} + T_{exe}(\tau)$, the surely running range of a task $\tau$ is given as the time range between $t_{lst}(\tau)$ and $t_{est}(\tau) + T_{exe}(\tau)$, otherwise the surely running range is empty. Note that $t_{est}(\tau)$ and $t_{lst}(\tau)$ denotes $t_{est}(\tau, p)$ and $t_{lst}(\tau, p)$ in any path $p$ in a valid process, because $t_{est}(\tau, p)$ is same as $t_{est}(\tau, p')$ in a valid process regardless of paths $p$ and $p'$. The runnable range of a task $\tau$ is given as the time range between $t_{est}(\tau)$ and $t_{lst}(\tau) + T_{exe}(\tau)$ as shown in the figure. Finally, given two



(a) A path  (b) Runnable time chart  (c) Runnable time range
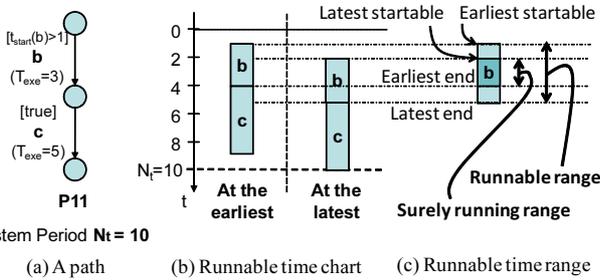
Figure 4: Runnable time range

task $\tau_1$ and $\tau_2$, they should be assigned into a same context if the surely running range of $\tau_1$ and that of $\tau_2$ overlap. Also, they can be assigned into different contexts if the runnable range of $\tau_1$ and that of $\tau_2$ do not overlap.

In the proposed algorithm, we construct a graph where a node represents a task in a valid system and an edge between two tasks represents a relation whether they should/can be assigned into a same context. We call this graph a *task dependency graph*. We give a single edge between any two tasks the runnable range of which overlaps each other. We also give a double edge between any two tasks the surely running range of which overlaps each other. However, we do not give any edge between two tasks if at least one of the following conditions satisfies:

> The tasks are in the same process and are not in the same path, and

All the paths which contain one of the task are not synchronized with all the paths which contain the other task.

For example, Fig. 5 shows the runnable time range and the task dependency graph of the system in Fig. 1.



(a) Runnable Time Range



(b) Task Dependency Graph
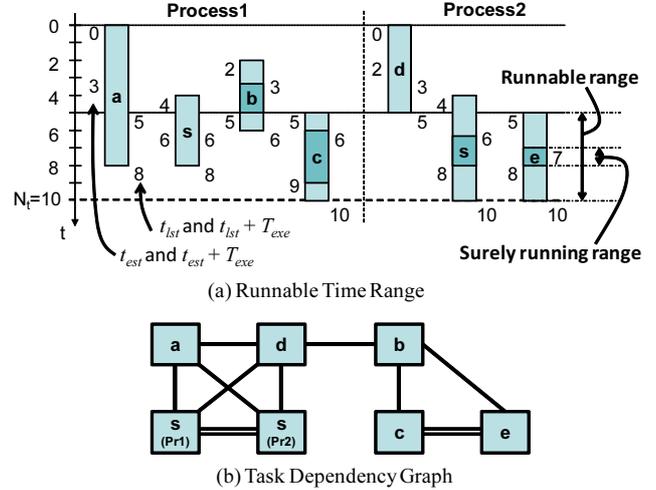
Figure 5: Example of `System1`

#### 4.2.3 Assignment Tasks into Contexts using Task Dependency Graph

In this phase, we decompose the task dependency graph of a given system into several smaller graphs of which the maximum size is as small as possible. If the number of derived graphs is larger than the number of contexts $N_c$, the graphs are merged into $N_c$ graphs. Finally, each the graph corresponds to each context of a DRP.

Tasks connected by double edges in a task dependency graph should be assigned into a same context of a DRP, because they have run all together for at least one clock. Thus, we decompose a task dependency graph keeping the size of the maximum connected graph with double edges small.

In a task dependency graph, a single edge can be changed to a double edge or no edge due to the scheduling of the edge's both end tasks. That is calculated by updating the runnable time range of each tasks. Figures 6 and 7 show how to decompose a connected tasks.

Figure 6 shows an example of decomposing tasks connected by double edges. By assign Task b into two contexts redundantly, the size of the maximum connected tasks by double edges can be reduced although the total size of implemented tasks into contexts becomes larger. However, tasks in a clique shaped by double edges cannot be decomposed.

Figure 7 shows an example of decomposing tasks connected by single edges. Tasks connected by single edges can be assigned into a same contexts. It is also important to decompose tasks connected by single edges and make the size of the maximum connected tasks by single edges smaller. However, as shown in Fig. 7, after the decomposition, the single edge between Task a and b has been changed to a double edge. Thus, it needs to be carefully chosen how a connected task is decomposed.

(a) Before decomposing



(b) After decomposing
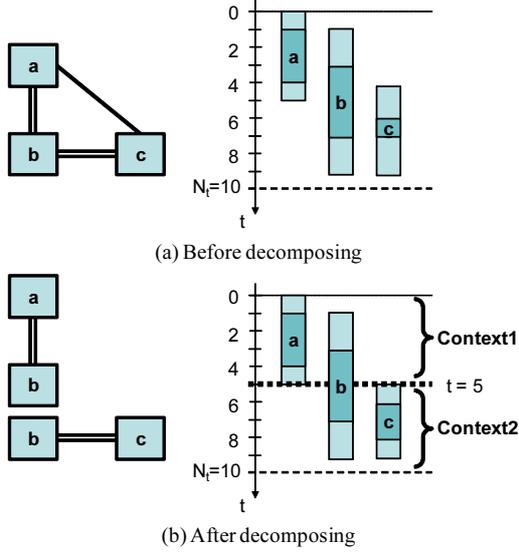
Figure 6: Example of Decomposing Tasks connected by Double Edges



(a) Before decomposing



(b) After decomposing
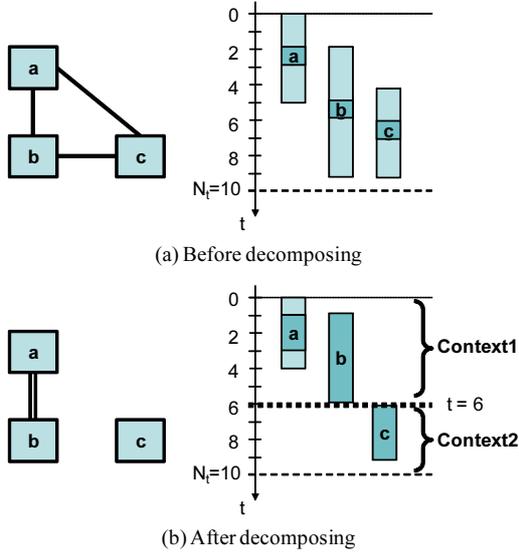
Figure 7: Example of Decomposing Tasks connected by Single Edges

Finally, after an assignment into contexts, the size of the maximum context is at least $\frac{S}{N_c}$. Here, $S$ denotes the sum of the tasks in a valid system, calculated as $S = \sum_\tau size(\tau)$ and note that $N_c$ is the number of contexts of a DRP. Therefore, in our algorithm, we decompose connected tasks in the following order of priority:

1. the maximum connected tasks by only double edges and the size of them is larger than $\frac{S}{N_c}$,

2. the maximum connected tasks by double edges and single edges and the size of them is larger than $\frac{S}{N_c}$,

3. the maximum connected tasks by only double edges, or

4. the maximum connected tasks by double edges and single edges.

A connected tasks is decomposed as follows.

Step 1. $C$ denotes a set of contexts. $C$ is initialized to the set of all tasks.

Step 2. Select one context $c$ from $C$ by the priority as mentioned above. $C \leftarrow C/\{c\}$.

Step 3. Sort the tasks in $c$ by $t_{est}$.

Step 4. Find a clock $t$ when the tasks are divided into two groups $G_1$ and $G_2$. $G_1$ consists of tasks with $t_{est} \leq t$ and $G_2$ consists of tasks with $t_{est} > t$ so that the total size of each group is (almost) the same.

Step 5. Find the cut set between $G_1$ and $G_2$.

Step 6. Given a double edge in the cut set and a task in $G_1$ is an end of the edge, the task is assigned to two new contexts $c_1$ and $c_2$. Other tasks in $G_1$ are assigned to $c_1$ and the tasks in $G_2$ are assigned into $c_2$.

Step 7. $C \leftarrow C \cup \{c_1, c_2\}$.

Step 8. Update the runnable time range of all the task.

Step 9. If all connected tasks in $C$ are clique then end, else goto Step 2.

We call a length between $t_{est}$ and $t_{lst}$ a *deadline slack*. When a graph is decomposed to two subgraphs, it is important to allocate the deadline slack to the subgraphs. In this algorithm, we give half of the deadline slack to each of the subgraphs. Later, we evaluate how the deadline slack is allocated into two subgraphs.

After the decomposition, we have derived the sets of tasks corresponding assignment into contexts. However, the number of the sets $C$ can be larger than $N_c$. Thus, we merge the sets into $N_c$ sets. In the proposed method, the smallest set of $C$ and the second-smallest set are merged until the number of sets becomes $N_c$.

## 5   Evaluation

We have compared the performance of the proposed heuristic algorithm. We implemented the algorithm in Java 5.0 on a Core2 Duo 1.2GHz CPU, 2GB RAM with Microsoft Windows XP Professional. We consider systems with $m$ concurrent processes, each of which consists of $n$ paths. For each $< m, n >$, we randomly derive 20 systems. In the following experiments, we have considered systems with the time period $N_t = 30$ and DRPs with four contexts ($N_c = 4$).

First, we evaluated how the deadline slack is divided into two subgraphs at a ratio of. Table 1 shows the summary of the obtained results for cases $5 : 5$, $6 : 4$, $7 : 3$, and $8 : 2$. This results shows that it is the best for a deadline slack to be divided into two subgraphs at a ratio of $5 : 5$.

Next, we have compared the performance of the proposed heuristic algorithm with a general ILP solver w.r.t. the execution time necessary for obtaining a task assignment and w.r.t. the quality of the obtained solutions. We used the ILP solver GLPK[20] in our experiments. For each derived system, we have derived its corresponding linear constraints, applied the proposed algorithm and the ILP solver, and found assignments. Then we calculate the average execution time of the algorithm and the ILP solver.Table 2 shows the summary of the obtained results for cases $< 2, 2 >$ and $< 2, 4 >$.

Table 1: Maximum Context Size against Ratio of Deadline Slack

| $< m, n >$ | $< 2, 2 >$ | $< 4, 4 >$ | $< 8, 8 >$ |
|---|---|---|---|
| $5 : 5$ | 43.9 | 84.8 | 128.7 |
| $6 : 4$ | 45.6 | 85.4 | 131.6 |
| $7 : 3$ | 47.1 | 100.6 | 130.8 |
| $8 : 2$ | 62.3 | 112.0 | 180.3 |

Table 2: Two Solver Results and Execution Time

| $< m, n >$ | $< 2, 2 >$ | | $< 2, 4 >$ | |
|---|---|---|---|---|
| | size | time | size | time |
| ILP | 26.7 | 0.16 sec. | 35.0 | > 10 h. |
| Proposed | 27.7 | 78 msec. | 39.0 | 125 msec. |

The results clearly show that the proposed algorithm outperforms the ILP in respect to execution time, and the difference becomes remarkable as the size of the system becomes bigger. For small systems $< 2, 2 >$, our method assigned the systems with the average logical area size 27.7 within 78ms and the ILP solver assigned it with the size 26.7 within 0.16sec. For larger systems $< 2, 4 >$, our method assigned the systems with the average logical area size 39.0 within 125ms and the ILP solver assigned it with the size 35.0 within 10hours.

## 6 Conclusion

In this paper, we have focused on a problem implementing a periodic concurrent system with real-time constraints into a multi-context DRP. We have proposed a heuristic algorithm to solve the problem efficiently. In the experimental results, we have confirmed that the proposed heuristic algorithm derives quasi-optical assignment results for large size examples in short time.

Considering the context switching overhead to our proposed algorithm and applying it to several types of real systems are our future work.

## REFERENCES

[1] B.P. Dave, and N.K. Jha, "CASPER: Concurrent Hardware-Software Co-Synthesis of Hard Real-Time Aperiodic and Periodic Specifications of Embedded System Architectures," *Proc. of DATE '98*, pp.118–125 (1998).

[2] T. Kitani, Y. Takamoto, K. Yasumoto, A. Nakata, and T. Higashino, "A Flexible and High-Reliable HW/SW Co-Design Method for Real-Time Embedded Systems," *Proc. of RTSS2004*, pp. 437–446 (2004).

[3] K. Ramamritham, "Allocation and Scheduling of Complex Periodic Tasks," *Proc. ICDCS '90*, pp. 108–115 (1990).

[4] D.-T. Peng, and K.G. Shin, "Assignment and Scheduling of Communication Periodic Tasks in Distributed Real-Time Systems," *Proc. ICDCS '89*, pp. 190–198 (1989).

[5] C.-J. Hou, and K.G. Shin, "Allocation of Periodic Task Modules with Precedence and Deadline Constraints in Distributed Real-Time Systems," *IEEE Trans. Comput.*, vol. 46, no. 12, pp. 1338–1356 (1997).

[6] R. Maestre, R. Hermida, F.J. Kurdahi, N. Bagherzadeh, and H. Singh, "Optimal vs. Heuristic Approaches to Context Scheduling for Multi-Context Reconfigurable Architectures," *Proc. ICCD'00*, pp. 575–576 (2000).

[7] K.M.G. Purna, and D. Bhatia, "Temporal Partitioning and Scheduling Data Flow Graphs for Reconfigurable Computers," *IEEE Trans. Comput.*, vol. 48, no. 6, pp. 579–590 (1999).

[8] B. Ouni, A. Mtibaa, and M. Abid, "Synthesis and Time Partitioning for Reconfigurable Systems," *Design Automation for Embedded Systems*, vol. 9, no.3, pp. 177–191 (2004).

[9] M. Kaul, and R. Vemuri, "Optimal Temporal Partitioning and Synthesis for Reconfigurable Architecture," *Proc. of DATE '98*, pp. 389–396 (1998).

[10] M. Sanchez-Elez, M. Fernandez, R. Maestre, R. Hermida, and N. Bagherzadeh, and F.J. Kurdahi, "A Complete Data Scheduler for Multi-Context Reconfigurable Architectures," *Proc. of DATE 02*, pp. 547–552 (2002).

[11] S. Banarjee, E. Bozorgzadeh, and N. Dutt, "Physically-Aware HW-SW Partitioning for Reconfigurable Architectures with Partial Dynamic Reconfiguration," *Proc. DAC '05*, pp. 335–340 (2005).

[12] S. Banarjee, E. Bozorgzadeh, and N. Dutt, "Considering run-time reconfiguration overhead in Task Graph Transformations for dynamically reconfigurable architectures," *Proc. FCCM '05* pp. 273–274 (2005).

[13] F. Redaelli, M.D. Santambrogio, and D. Sciuto, "Task Scheduling with Configuration Prefetching and Anti-Fragmentation techniques on Dynamically Reconfigurable Systems," *Proc. DATE '08*, pp. 519–522 (2008).

[14] ISO, "Information Processing System, Open Systems Interconnection LOTOS," *ISO 8807* (1989).

[15] H. Nakano, T. Shindo, T. Kazami, and M. Motomura, "Development of Dynamically Reconfigurable Processor LSI", *NEC Technical Journal*, Vol. 56, No. 4, pp. 99–102 (2003). http://www.necel.com/.

[16] T. Sunagawa, K. Ide, and T. Sato, "Dynamically Reconfigurable Processor Implemented with IPFlex's DAPDNA Technology", *IEICE Trans. Info. and Sys.*, Vol. E87-D, No. 8, pp. 1997-2003 (2004).

[17] H. Singh, M.-H. Lee, G. Lu, F.J. Kurdahi, N. Bagherzadeh, and E.M. Chaves Filho, "MorphoSys: An Integrated Reconfigurable System for Data-Parallel and Computation-Intensive Applications," *IEEE Trans. Comput.*, vol. 49, no. 5, pp. 465–481 (2000).

[18] S.C. Goldstein, H. Schmit, M. Budiu, S. Cadambi, M. Moe, and R. Taylor, "PipeRench: A Reconfigurable Architecture and Compiler," *IEEE Trans. Comput.*, vol. 33, no. 4, pp. 70–77 (2000).

[19] B. Mei, A. Lambrechts, D. Verkest, J.-Y. Mignolet, and R. Lauwereins, "Architecture Exploration for a Reconfigurable Architecture Template," *IEEE Trans. Design & Test of Comp.*, vol. 22. Issue 2, pp. 90–101 (2005).

[20] GNU Project, "GLPK (GNU Linear Programming Kit)", http://www.gnu.org/software/glpk/.

# Burstiness of Output Traffic by a Split Connection TCP Implementation

Toshihiro Shikama[†]

[†]Fukui University of Technology, Japan
shikama@fukui-ut.ac.jp

*Abstract* - This paper studies a TCP split connection implementation that employs separate TCP connections for a wired section and a wireless section, where a node connecting both sections forwards data from one section to another. We assume that transmission errors occur in the wireless section. From the results by the simulations using ns-2, this paper shows that the implementation produces large forward data due to losses of packets over the wireless section. It also shows that the output traffic from the forwarding node becomes bursty due to the large forward data. The effects of the bursts of packets are evaluated for the case where the output traffic from the forwarding node is fed into a node connected to a bottleneck link in the wired section. We found that a large number of packets are queued even when a packet loss rate is relatively small. The problem of bursts by the forwarding node has not been identified yet and its mitigation is left for further study.

*Keywords*: PEP, TCP Proxy, Split Connection, Bursty Traffic, 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, the congestion control is performed unnecessarily; the throughput of TCP is lowered.

One approach to mitigate this problem is to improve the congestion control of TCP itself. A typical version of TCP is Westwood [1][2], 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.

Another approach is to employ a PEP (Performance Enhancing Proxy) that terminates a TCP connection from a source terminal and establishes another TCP connection to the destination terminal [3]. It forwards receive data from one TCP connection to another one. As the large round-trip delay causes low throughput and the PEP makes the delay short by splitting the TCP connection, the throughput is improved. 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 [4].

Although the termination of a TCP connection by the PEP has a significant impact on the characteristics of IP flow, a study on this aspect has not been done so far. This paper focuses on the property of output traffic produced by the PEP and identifies the problem concerning bursts of packets caused by the termination of a TCP connection at the PEP.

The rest of this paper is organized as follows: Section 2 explains the architecture of the split TCP connection and its problem regarding output traffic. Section 3 describes the simulation model and associated assumptions. Section 4 describes simulation results and discussion. Finally our conclusion is presented in Section 6.

## 2   Output Traffic by a Split TCP Connection Implementation

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 part. We also call the wired network part "a wired section". As the PEP splits the round trip delay of an end-to-end TCP connection, an error recovery of a lost segment over the wireless section is done quickly. This leads to the improvement of the total TCP throughput.

Figure 1 illustrates an example of 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 segment, 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". 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 immediately issued by the PEP after the arrival of the retransmitted segment. The burst of packets might be harmful, since it may cause buffer overflows at nodes forwarding these packets on the wired section.

With regard to this busty packets from TCP, RFC 2581 specifies that TCP should employ the slow start after a silent period of more than one RTO (Retransmission Time Out) [5]. In the normal case, when an application program issues a send request of a large amount of data, the transmission of seg-
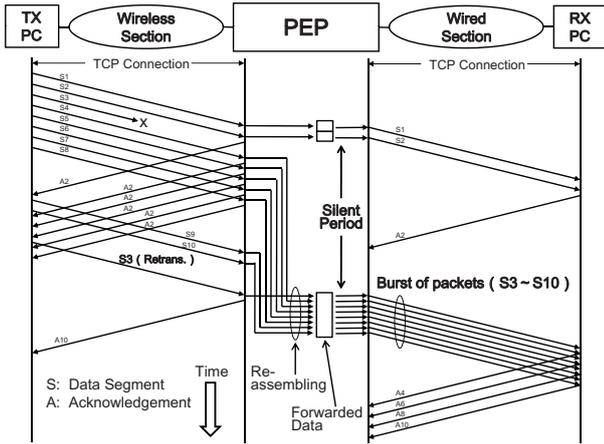
Figure 1: Data forwarding by a PEP and its output traffic

ments begins with the slow start and reaches to the self clocking state after some transient period [6]. The burst of packets occurs during the period of the slow starts only and its effect is considered to be limited.

In Figure 1, there is a silent period over the wired section from the loss of the packet to the reception of the retransmitted packet. However the duration of this period is comparable with one round-trip time of the wireless section, which is much less than RTO period. Therefore the slow-start is probably not invoked in this case; the burst of packets derived from the large forwarding data is likely fed into the wired section.

## 3 Simulation Models

This paper evaluates output traffic from the PEP by simulations using ns-2 [7]. Figure 2 shows the simulation model for the case where the PEP is 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 date 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 date to the upper application layer, we added this function to TCP sink in ns-2. This is indicated as "TCP Sink+" in Figure 2. We measured the distribution of the sizes of forward data inside the PEP and the number of packets included in a burst sent by the PEP.

Figure 3 represents the relationship between forward data inside the PEP and bursts of packets issued by the PEP. Since TCP provides the byte oriented streaming service, TCP segments received by the PEP are reassembled to variable length data, which is forwarded to the next TCP connection. The next TCP that accepts the data divides it into segments. These generated segments are sent over the wired section as packets. A series of packets sent back-to-back at the rate of the interface to node n2 is called a burst. Since the number of packets in a burst is affected by the size of the congestion window of TCP and the number of outstanding segments, this number is not necessarily consistent with the size of the forward data.

The bandwidth of the wireless section is assumed to be 5 Mbit/s, while the PEP is connected to node n2 with the interface rate of 100 Mbit/s and node n2 is connected to destina-
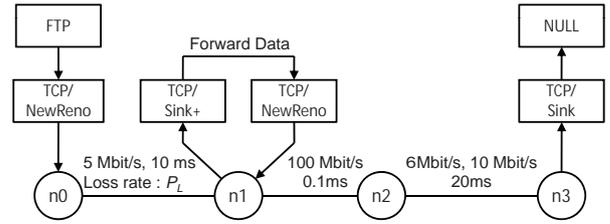


Figure 2: Simulation model for the case with PEP
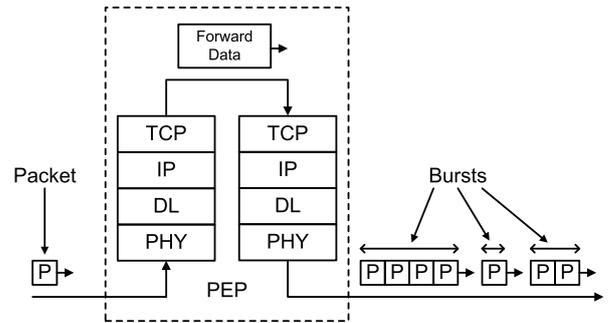


Figure 3: Relation between forward data and bursts of packets
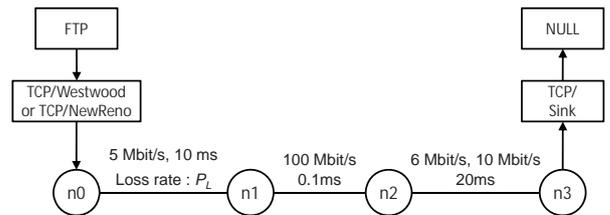


Figure 4: Simulation model for the case without PEP

tion node n3 with the rate of 6 Mbit/s or 10 Mbit/s, which are assumed to be the bottleneck bandwidth over the wired section. Since there is a difference of packet transmission rate between input and output links in node n2, bursts of packets might be queued in this node. We measured the number of packets in a burst on the 100 Mbit/s interface between nodes n1 and n2.

In order to confirm the improvement of throughput by the PEP we also simulate the case where a PEP is not employed. Figure 4 shows the model of no PEP, where either NewReno TCP or Westwood TCP is employed end-to-end for comparison. Table 1 summarizes the simulation parameters.

## 4 Simulation Results and Discussion

### 4.1 Throughput

Figure 5 shows the relation between the packet loss rate and the throughput. In this figure NOPEP (W) represents the no PEP case where Westwood TCP is employed end-to-end, while NOPEP (N) represents the no PEP case where NewReno TCP is employed end-to-end. Throughput generally decreases as the packet loss rate becomes large. However, the decrease of throughput is suppressed significantly by the PEP. In the cases of no PEP, where an error recovery is per-

Table 1: Simulation parameters

| | |
|---|---|
| Bandwidth of the wireless section $B_R$ | 5 Mbit/s |
| Delay of the wireless section | 10 ms |
| TCP for the wireless section | NewReno, Westwood |
| TCP window size for the wireless section | 128 KB |
| Generation of packet losses | random |
| Packet loss rate $P_L$ | 0.0001, 0.0002, 0.0005, 0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1 |
| Bandwidth of the wired section $B_W$ | 6 Mbit/s, 10 Mbit/s |
| TCP for the wired section | NewReno |
| TCP window size for the wired section | 256 KB |
| Delay of the wireless section | 20 ms |
| The number of TCP connections $N_{TCP}$ | 1, 5 |
| Interface rate from the PEP to the wired section | 100 Mbit/s |
| The number of buffers at each node | unlimited |
| Simulation time | 1500 sec |
| The number of simulation runs | 12 |
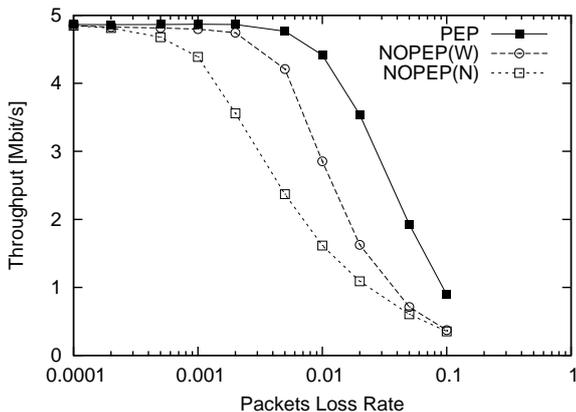


Figure 5: Packet loss rate vs. throughput

formed end-to-end, it is clear that Westwood TCP improves the throughput compared with NewReno TCP.

## 4.2 The distributions of the forward data sizes and the number of packets in a burst

Figure 6 shows the distributions of both the forward data sizes and the number of packets in a burst, where the number of TCP connections $N_{TCP}$ is 1 and the packet loss rate $P_L$ is changed to 0.0001, 0.001, 0.01, and 0.1. 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 bytes) to compare the burst size. In this figure rectangle boxes represent the distribution of the forward data sizes, while black dots show the distribution of the number of packets in a burst.

The maximum size of forward data is about 90, which is consistent with the TCP window size (128 kB) of the wireless section. We can observe the trend that the distribution of the forward data size and the number of packets in a burst are well consistent. This means that the PEP generates bursts of

packets corresponding to the forward data sizes. In the cases of small packet loss rates, although the frequency of bursts is generally small, there is a trend that large size bursts are generated.

Figure 7 shows the same distributions as Figure 6, where the number of TCP connections $N_{TCP}$ is changed to 5. Similar trends are observed as the case of the single TCP connection. However, when the packet loss rate is small ($P_L = 0.0001, 0.001$), the distribution of the number of packets in a burst is different from that of the forward data sizes. Frequency of bursts of which size is larger than 1 becomes small as compared with the distribution of the forward data sizes. The reason is as follows: as the number of TCP connections increases, the number of outstanding segments waiting for ACKs on the wired section becomes large. In this case, a new segment can be sent only after a new ACK arrives and the window is updated and opened by the ACK. This means that the self clocking state likely continues in the cases of small packet loss rates.

## 4.3 The distribution of inter-arrival times of forward data

Figure 8 shows the distribution of inter-arrival times of forward data in the case where the number of TCP connections $N_{TCP}$ is 1 and the packet loss rate $P_L$ is 0.001. In this small packet loss case, the inter-arrival times take the values of less than 200ms. Since the minimum value of the RTO time is 200ms and retransmissions by the RTO timer needs more than 200ms, the distribution shows that the recovery of lost segments is done by the fast retransmit and not by a timeout.

## 4.4 The correlation between the inter-arrival times and the forward data sizes

It can be expected that a strong correlation may exist between the sizes and inter-arrival times of forward data. Figure
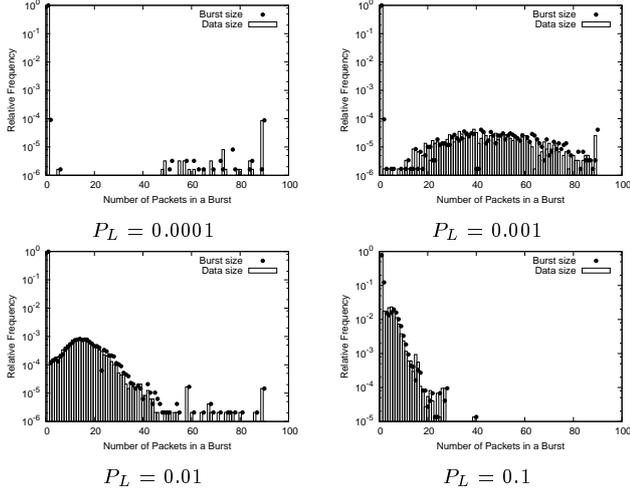
Figure 6: Distribution of the forward data sizes and the number of packets in a burst, where $N_{TCP}$ is 1.
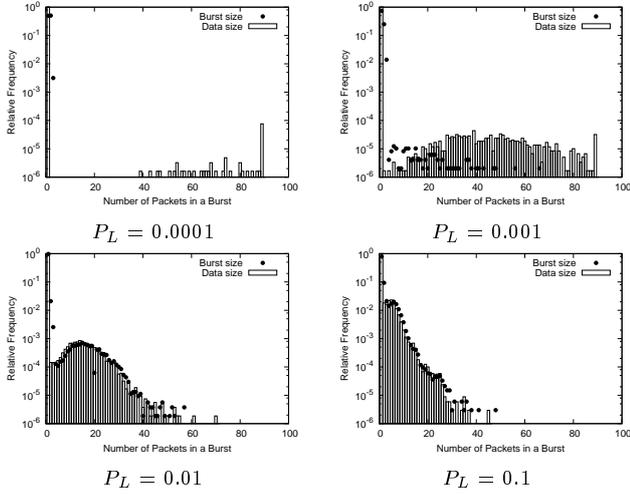


Figure 7: Distribution of the forward data sizes and the number of packets in a burst, where $N_{TCP}$ is 5.
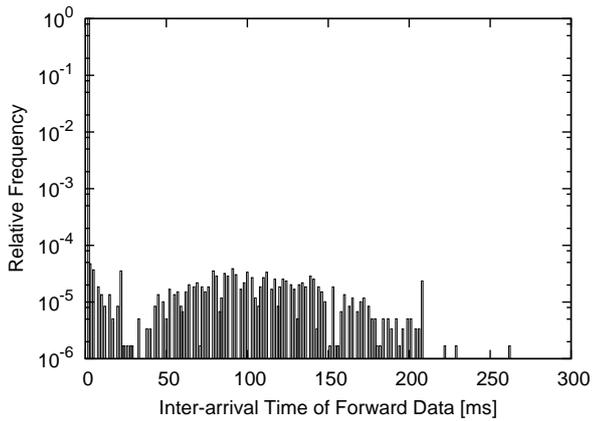


Figure 8: Distribution of inter-arrival times of forward data, where $P_L$ is 0.001 and $N_{TCP}$ is 1.

9 shows the scatter diagram of the pair of these two values, where the number of TCP connections $N_{TCP}$ is 1 and the packet loss rate $P_L$ is 0.001. We can observe the strong correlation between the pair of two values; as the inter-arrival time

of forward data becomes large, the forward data size also increases. In this condition, a lost segment is mainly recovered by the fast retransmit; the number of segments received correctly after a lost segment is approximately proportional to the interval needed to recover the lost segment. This is considered to be the reason why the strong correlation exists between the sizes and inter-arrival times of forward data.

Figure 10 also shows the same scatter diagram as Figure 9, except that the packet loss rate $P_L$ is 0.01. When the inter-arrival times are less than 120 ms, the strong correlation can be observed, however the dispersion of plotted data is larger than the case of Figure 9. The correlation cannot be observed for the inter-arrival times larger than 220ms. This is the effect of retransmissions by timeouts. When the packet loss rate becomes large, the lost segments cannot be recovered by the fast retransmit only, the possibility of retransmissions by timeouts increases.

Figure 11 shows the time changes of cwnd, where the packet loss rate $P_L$ is 0.001 and the number of TCP connections $N_{TCP}$ is 1. This figure indicates that the fast retransmit is employed for retransmissions. Figure 12 shows the same cwnd changes, where the packet loss rate $P_L$ is 0.01. In this figure, retransmissions by both the fast retransmit and timeout are observed. The observation of these figures of cwnd is consistent with the scatter diagrams shown in Figures 9 and 10.

Figure 13 shows the scatter diagram of the same condition as Figure 9, except that the number of TCP connections $N_{TCP}$ is changed to 5. Since an inter-arrival time of forward data on one TCP connection is fragmented by the arrivals of forward data from other 4 TCP connections, the correlation between the sizes and inter-arrival times of forward data becomes small. However, plotted circles tend to be distributed around the upper side of the correlation line observed in Figure 9 because of the fragmentation of an inter-arrival time.

The queueing analysis for bursty traffic has been studied such as $M^{[X]}/G/1$ and $G^{[X]}/G/1$ [8]. It is expected that we can apply such results to evaluate queueing effects of the bursty traffic at a node forwarding the traffic. However, the analysis generally assumes that the inter-arrival times and the burst sizes are independent (no correlation). The strong correlation observed in Figures 9 and 10 indicates that we cannot apply the analytical approach to evaluate queueing effects at the forwarding node.

## 4.5 The distribution of queue length at the bottleneck node in the wired section

The effect of the bursts of packets is typically observed as large queueing at a bottleneck node forwarding the bursts. Figures 14 and 15 show the distributions of queue length at the bottleneck node n2, where the number of TCP connections $N_{TCP}$ is 1 and 5, respectively. Each figure shows the cases where the packet loss rate $P_L$ is changed to 0.0001, 0.001, 0.01, and 0.1. In Figure 14, although the bandwidth of the wired section $B_W$ is 1.2 times larger than that of the wireless section $B_R$, the large queue length is observed even in the case where the packet loss rate $P_L$ is small (0.0001 and 0.001). When the number of TCP connections becomes 5, the maximum queue length becomes slightly larger compared
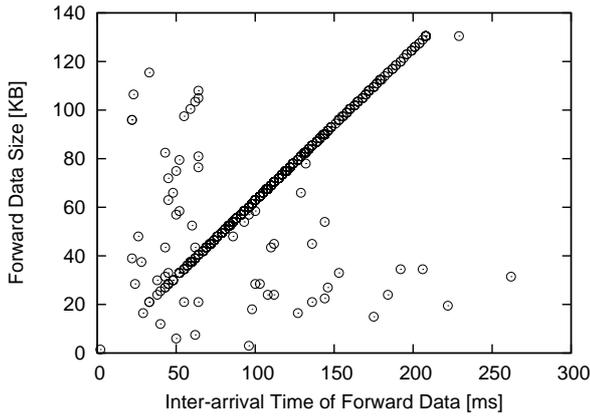
Figure 9: Scatter diagram of the sizes and inter-arrival times of forward data, where $P_L$ is 0.001 and $N_{TCP}$ is 1.
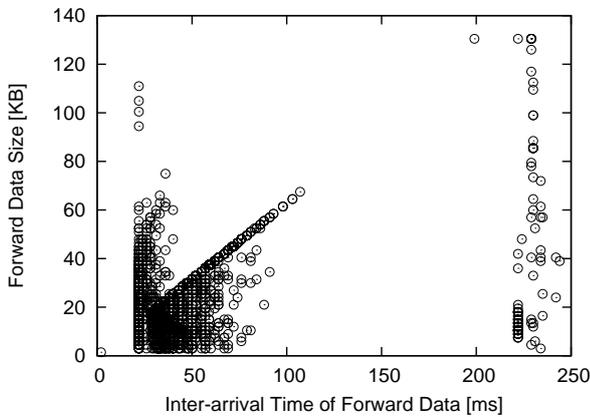


Figure 10: Scatter diagram of the sizes and inter-arrival times of forward data, where $P_L$ is 0.01 and $N_{TCP}$ is 1.

with the cases of the single TCP connection. However, there is not significant difference on this value.

## 4.6 The average of queue length at the bottleneck node in the wired section

Figures 16 and 17 show the average queue length and its standard deviation, where the bandwidth of the wired section $B_W$ is 6 Mbit/s and the number of TCP connections $N_{TCP}$ is 1 and 5, respectively. These figures also indicate 95% confidence interval for each plotted value.

When the number of TCP connections $N_{TCP}$ is 1, the average queue length takes the maximum value at around the packet loss rates of 0.001 and 0.002. The standard deviation also becomes the largest when the packet loss rate is around 0.001, and its value is about 2 times larger than the average value. This means that the variation of the queue length is large. When the number of TCP connections $N_{TCP}$ is 5, the same trend is observed, but the standard deviation is relatively small compared with the average value.

Figures 18 and 19 show the same values as Figures 16 and 17, where the bandwidth of the wired section $B_W$ is increased to 10 Mbit/s. When we compare Figures 18 and Figures 16 where the number of TCP connections is 1, the average queue
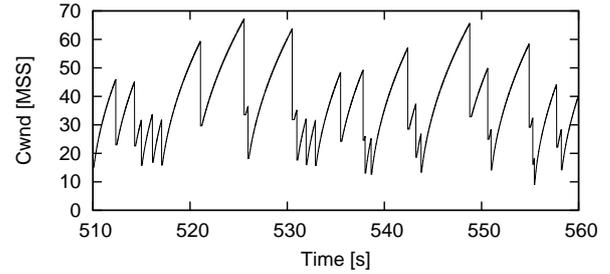


Figure 11: Example of cwnd change, where $P_L$ is 0.001 and $N_{TCP}$ is 1.
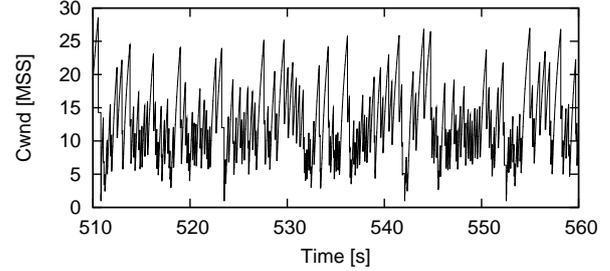


Figure 12: Example of cwnd change, where $P_L$ is 0.01 and $N_{TCP}$ is 1.
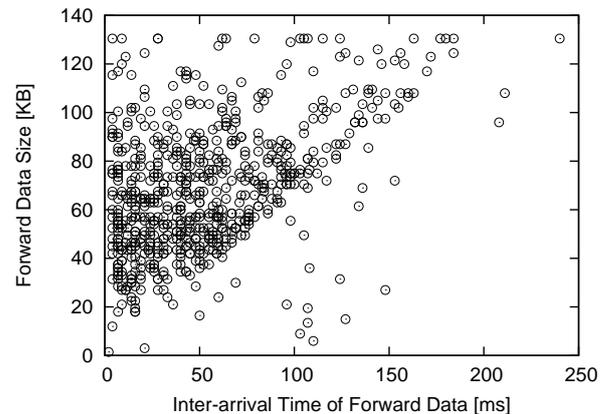


Figure 13: Scatter diagram of the sizes and inter-arrival times of forward data, where $P_L$ is 0.01 and $N_{TCP}$ is 1.

length in Figure 18 is decreased to about a quarter of the values in Figures 16, while the standard deviation is decreased to a half. Although the average queue length becomes small, the variation of queue length relative to the average value tends to increase.

If the PEP is not employed and the error recoveries are performed end-to-end, the queueing at the bottleneck node never occurs. Although the results of no PEP case are not plotted in the figures, the queue length always takes value 0. It is clear that a split TCP implementation causes significant queueing at a bottleneck node because of the bursts of packets issued by the PEP. Mitigation of this problem is left for further study.
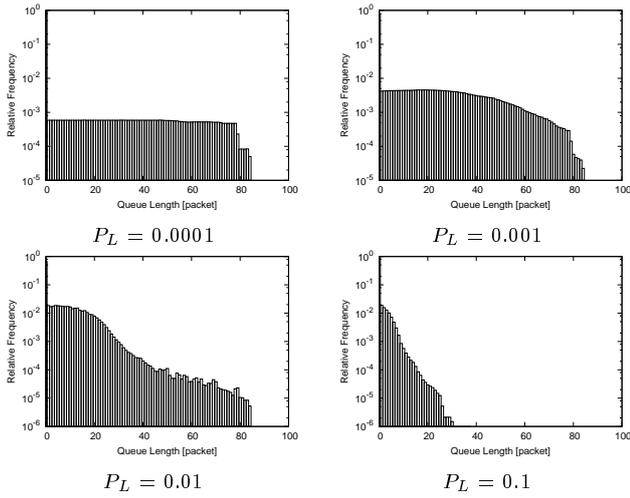
Figure 14: Distribution of the queue length, where $N_{TCP}$ is 1 and $B_W$ is= 6 Mbit/s.
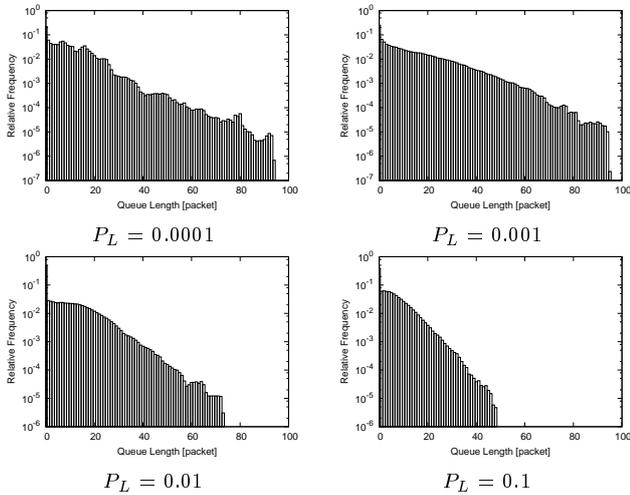


Figure 15: Distribution of the queue length, where $N_{TCP}$ is 5 and $B_W$ is 6 Mbit/s.
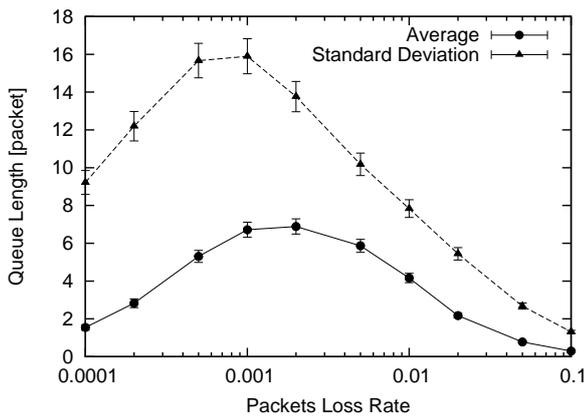


Figure 16: Average and standard deviation of queue length, where $N_{TCP}$ is 1 and $B_W$ is 6 Mbit/s.

## 5 Possible approaches to mitigate the burstiness of the PEP output

As described in section 4, the output traffic by the PEP becomes bursty in cases where packets are lost by transmission
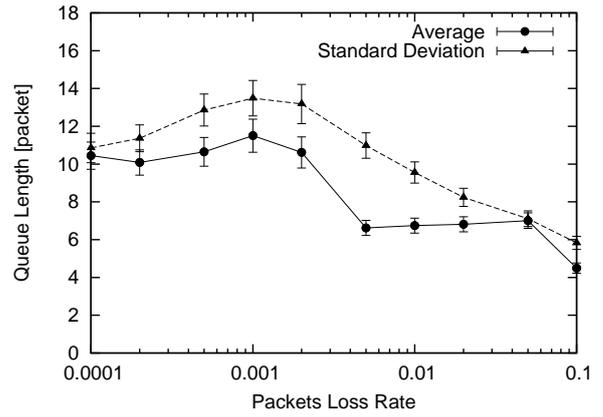


Figure 17: Average and standard deviation of queue length, where $N_{TCP}$ is 5 and $B_W$ is 6 Mbit/s.
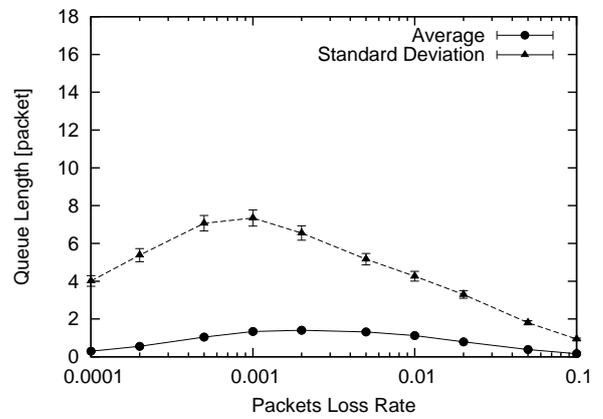


Figure 18: Average and standard deviation of queue length, where $N_{TCP}$ is 1 and $B_W$ is 10 Mbit/s.
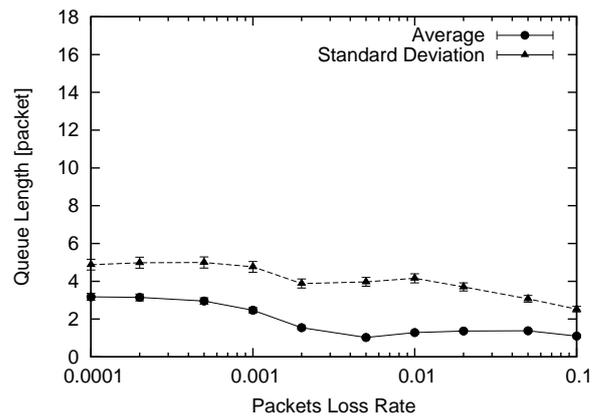


Figure 19: Average and standard deviation of queue length, where $N_{TCP}$ is 5 and $B_W$ is 10 Mbit/s.

errors. A further study is needed to mitigate this problem. Our goal is not only to reduce the burst sizes on the output link but also to suppress the large queueing at the bottleneck node in the wired section. There are three possible approaches considered.

The first approach is to utilize the "maxburst" parameter. There are TCP implementation and its variant supporting the

"maxburst" parameter, which limits the maximum number of packets sent back-to-back on the output link when a transmission of a large size data is requested while the size of the transmission window is large enough [9], [10]. The second approach is to introduce a kind of traffic shaping between two TCP connections. The third approach is to change the PEP architecture, where an out-of-order segment is forwarded without being reassembled by the PEP.

The first approach is simple, but its effectiveness on suppressing the large queueing at the bottleneck link is questionable. Although the second approach is straightforward, it hard to define the suitable shaping rate. In this approach we also have to consider the coordination of the TCP flow control between two TCP connections to avoid buffer overflows in the PEP. Since the TCP implementation in current ns-2 does not include the dynamical change of the available window at the receiver, we have conducted the simulations on condition that throughput of the wired section is always larger than the wireless section. Further modification of ns-2 is required to evaluate the coordination. The third approach is interesting since bursts of packets never occur. However it seems that the implementation becomes very complex and hard.

We are now performing studies and performance evaluations concerning these approaches.

## 6 Conclusion

This paper has investigated the architecture of the split connection TCP implementation. We have focused on the output traffic where the PEP that forwards receive data from one TCP connection on the wireless section to another TCP connection on the wired section. When losses of packets are caused by transmission errors, we have observed that the output traffic from the PEP becomes bursty due to the reassembling function done by the receiving side of TCP covering the wireless section. We also have found that the burstiness becomes significant at packet loss rates that are relatively small, and the effects of the bursts become small as the packet loss rate increases. We have observed the strong correlation between the inter arrival times of bursts and burst sizes. This means that analytical approaches cannot be applied to evaluate the queueing effects at a node forwarding the bursty traffic. 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. We have found that a large number of packets are queued when the packet loss rate is relatively small.

All these observations concerning the PEP have not been identified yet. As the rate of wireless links is increasing, the PEP architecture will be important to attain high TCP throughput. Accordingly, effects of the bursty traffic by the PEP become also significant. Further studies and evaluations are needed to mitigate this problem.

### Acknowledgment

## REFERENCES

[1] 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.

[2] L. A. Grieco and S. Mascolo, "Performance evaluation and comparison of Westwood+, New Reno and Vegas TCP congestion control, " ACM Computer Communication Review, Vol. 34(2) , April 2004.

[3] J. Border, M. Kojo, J. Griner, G. Montenegro, and Z. Shelby, "Performance Enhancing Proxies Intended to Mitigate Link-Related Degradations, " IETF, RFC3135, June 2001.

[4] M. Meyer, J. Sachs, and M. Holzke, "Performance evaluation of a TCP proxy in WCDMA networks," IEEE Wireless Communications, vol.10, no.5, pp.70-79, Oct. 2003.

[5] M. Allman, V. Paxson, and W. Tevens, "TCP Congestion Control, " RFC 2581, April 1999.

[6] W. R. Stevens, "TCP/IP Illustrated, Volume 1: The Protocols, " Addison-Wesley, 1994.

[7] Network Simulator - ns (version 2), http://www.isi.edu/nsnam/ns/, 2009.

[8] L. Kleinrock, "Queueing Systems, vol. 2," John Wiley and Sons, 1976.

[9] J. Iyengar, E. Blanton, and M. Allman, "TCP Burst Mitigation Through Congestion Window Limiting," draft-iyengar-burst-mitigation-00.txt, Nov. 2005.

[10] R. Stewart, L. Ong, I. Arias-Rodriguez, K. Poon, A. Caro, and M. Tuexen, "Stream Control Transmission Protocol (SCTP) Specification Errata and Issues," RFC 4460, April 2006.

# A Method of Selecting Optimal Measures for Security and Usability with Fault Tree Analysis and State Transition Diagram

Koichi Kato[*] and Yoshimi Teshigawara[**]

[*, **]Graduate School of Engineering, Soka University, Japan
[*]kokatou@soka.ac.jp, [**]teshiga@t.soka.ac.jp

**Abstract** -A network must have enough usability to achieve its constructional purpose and enough security to protect information assets. Security and usability, however, are in a trade-off relation. To select appropriate measures is difficult because of following reasons: (i) chain relations exist in risks, in services use and between them, and (ii) risks, usability and measures have complex relations. This paper proposes a method of analyzing risk/usability in consideration of chain relations and selecting measures for security and usability, using fault tree analysis (FTA) and state transition diagram (STD). Furthermore, by using probability of state transition, the STD can be converted to a Bayesian network and thereby our method can infer causes of incidents. As a result, the method can also select effecting measures to manage and monitor the network.

**Keywords**: Risk Management, Usability, Fault Tree Analysis, State Transition Diagram, Bayesian Network.

## 1 INTRODUCTION

Recently, a variety of network environments are being constructed, such as home networks, enterprise or university networks and high-speed public wireless networks. These networks are established for various purposes of accessing to the Internet as well as resource sharing, business efficiency, etc.

On the other hand, information systems on a network have many risks such as hacking and information leakage. The occurrence of a risk event may cause not only direct damages by business suspension or system recovery works but also loss of organizational credibility. Therefore, today a variety of measures to reduce risks are taken in a network.

In order to achieve original purposes of a network, striking a good balance between security for protecting information assets and usability of services is important. However, security and usability are generally in a trade-off relation. For example, security measures can disturb comfortable use of information systems by increasing steps of the using process for users or slowing down the execution speed of the systems. Excessive security measures decrease usability level. In contrast, excessive usability decreases security level. Consequently, to balance security with usability is difficult.

In addition, to manage and monitor the points related to risks and usability are important measures to check whether security issues happen and services are properly provided [1]. However, to decide appropriate monitoring points is difficult in a network locating with many devices.

As the related work, in order to decide the priority of risks to deal with, Zuccato [2] described the decision matrix and Guan et al. [3] evaluated security with Analytic Hierarchy Process (AHP) and the risk level matrix. They analyzed risks from a certain viewpoint such as an expert and did not consider existence of multiple stakeholders. Yajima et al. [4] suggested the multiplex risk communicator to decide measures based on the agreement among stakeholders. However, they did not analyze usability in detail. Kotenko and Stepashikin [5] as well as Wang [6] described each security evaluation method using the attack graph. They could analyze events of incidents and decide measures visually in consideration of the network configuration. However, they cannot express effects on usability caused by the selected measures.

Our research has approached the matter in two ways: (i) a method of selecting optimal measures at changing implemented measures [7] and (ii) expression model of risks, usability and measures [8]. This paper proposes a method of selecting optimal measures to construct a network that has enough usability to operate businesses and enough security to protect information assets by combining above-mentioned method and model. This method analyzes phases of risks and services use with FTA and STD. As a result, we can quantify risks and usability and select optimal measures intuitively by visualizing risks, usability and measures.

## 2 RESEARCH ISSUE

### 2.1 Risk and Usability Analysis with Chain Relations

There exist relations of risk chains that the occurrence of a single risk event causes other multiple risks. In addition, in the risk chains, an elemental event in the process of an occurrence of a risk event may diverge to other risks. Similarly, there exist relations of usability chains that deterioration of a usability level causes a decrease in other usability levels. Therefore, analyzing relations among risks and usability with chain relations is needed to maintain usability levels and to reduce risk levels.

### 2.2 Relations among Risk, Usability and Measures

Relations among measures and risks/usability are complex. Several measures can be taken against a single risk. On the contrary, a measure also can be effective to some risks. Moreover, measures for risks may affect some
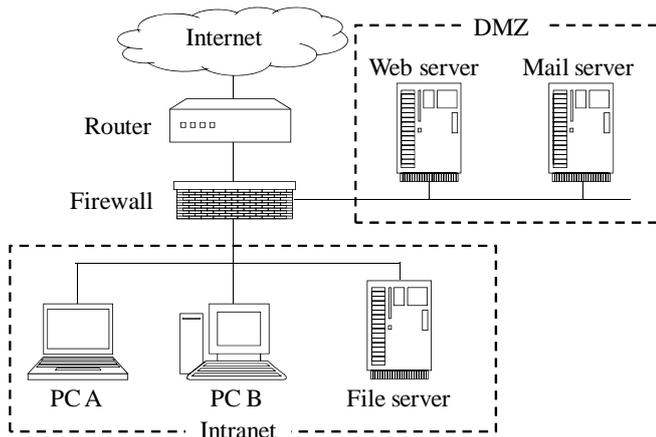
Figure 1: Example of network configuration.

usability. Therefore, analyzing these relations properly and effectively is needed.

## 2.3 Optimal Measures Selection

We define the meaning of "optimal measures" as a combination of measures that system administrators and users accept with satisfaction about security and usability at the time. Our research targets both phases of implementing measures and changing ones. The requirements for security and usability can vary depending on the situation such as that administrators and/or users require higher level of security or usability.

Objective evaluation of security and usability level earned by candidate measures are needed to select appropriate measures. Our method quantifies probability and the value of risks, which are general metric, and usability. The value of usability in our method is the rate of comfortable service use based on a completely free use of the service with no affects by measures.

We may select excessive or insufficient measures which cause unexpected security and usability decrease only with calculating theoretical optimal measures. Therefore, a scheme to decide appropriate measures intuitively for system administrators and users is needed.

In addition, the validity of monitoring points for security/usability cannot be evaluated because the points are often decided by experimental rules. Moreover, there exist no objective metrics to infer causes of incidents at the occurrence of a risk event. Then objects of analysis to identify the causes may spread unnecessarily. Therefore, a scheme to select monitoring points and specify the causes of incidents is needed.

## 3 RISK AND USABILITY ANALYSIS WITH FTA

Our method adopts FTA as a quantification method. FTA has following features: (i) it can analyze preventive factors against an achievement of a specific goal, (ii) it can organize measures to control each factor and (iii) it can decide appropriate measures to exact points. FTA makes Fault Tree (FT) that the top is an event as the result of an occurrence of
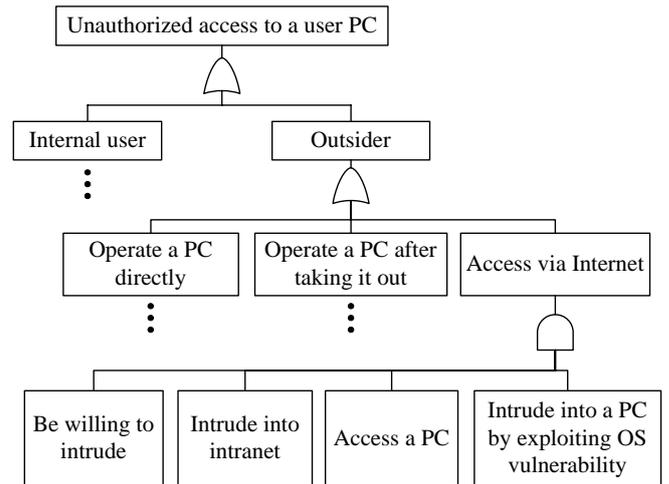


Figure 2: Fault tree for unauthorized access to a user PC.

Table 1: Analysis of risks and measures.

| basic event | probability | measure | risk decrease | state |
|---|---|---|---|---|
| Be willing to intrude | 0.7 | - | - | - |
| Intrude into intranet | 0.7 | Authentication for network access | 0.9 | 0 |
| | | FW access control | 0.7 | 1 |
| Access a PC | 0.7 | At a PC: PFW access control | 0.5 | 1 |
| Intrude into a PC by exploiting OS vulnerability | 0.4 | At a PC: OS update | 0.7 | 1 |

other causal events and these events are joined by logical AND/OR gates [9].

## 3.1 Network Example

To describe our method with some tangible examples, we assume a simple network example as shown in Figure 1. The network is separated into DMZ and intranet. A web server and a mail server are located on the DMZ. General users work using each user PC. The users can use files on a file server, browse web sites and send and receive e-mails.

## 3.2 Risk Analysis

Generally, some phases exist in the process of an occurrence of a risk event. A risk can be inhibited by measures preventing from occurring each phase based on the concept of defense in depth [10].

In order to quantify risk probabilities, our method makes FTs of risks. First, an unexpected risk event is placed at the top of the FT. Second, attack phases are developed at the bottom as shown in Figure 2. Third, measures related to each basic event, which is a leaf of the tree, are clarified. Finally, the probability of a basic event, the rate of risk decrease in an object basic event by a measure and the implementation state of a measure, which has yes or no selectively, are assigned as shown in Table 1. Note that the acronyms means firewall (FW) and personal firewall (PFW)
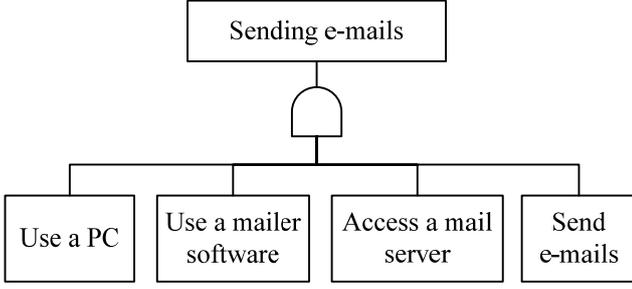
Figure 3: Fault tree for the usability of rending e-mails.

Table 2: Analysis of a service usage and measures.

| basic event | usability | measure | usability decrease | state |
|---|---|---|---|---|
| Use a PC | 1 | - | - | - |
| Use a mailer software | 1 | At a PC: password for a mailer software | 0.3 | 1 |
| Access a mail server | 1 | At a PC: PFW access control | 0.1 | 1 |
| Send e-mails | 1 | At a mail server: Authentication for sending e-mails | 0.1 | 1 |



Figure 4: STD of unauthorized access to a user PC.
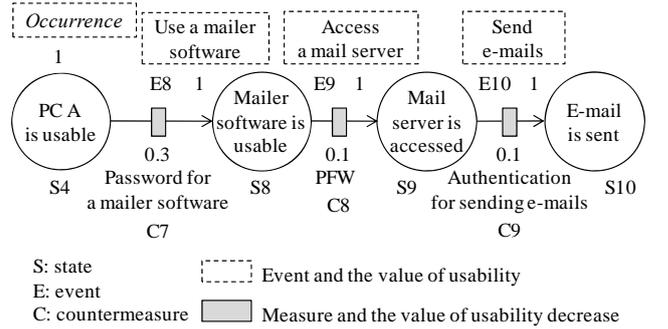


Figure 5: STD of sending e-mails.

and the value of the implement state (described as *state* in Table 1) has 0 as "not implement" or 1 as "implement".

Risk probability is formulated as follows. The probability of the top event is calculated based on minimal cut sets, which are minimum collections of basic events such that if they all occur the top event also occur.

The $P_{top}$, which is the probability of the top event, gives

$$P_{top} = 1 - \prod_{c \in C} \left\{ 1 - \prod_{e \in Ec} P_e \prod_i \left( 1 - X_i \Delta P_{e,i} \right) \right\} \qquad (1)$$

where $c$ is a minimal cut set, $C$ is a set of $c$, $e$ is a basic event in a cut set, $E_c$ is a set of $e$ in $c$, $P_e$ is the probability of $e$, $X_i \in \{0, 1\}$ is the implementation state of measure $i$ and $\Delta P_{e,i}$ is the rate of risk decrease in $e$ by $i$.

## 3.3 Usability Analysis

Some phases exist in the process of using a service as risks. However, service use may become insufficient because security measures may prevent from realizing some phases.

In order to quantify usability of services, our method makes FTs of services use. First, an object service is placed at the top of the FT. Second, using phases are developed at the bottom as Figure 3. Third, measures related to each basic event are clarified. Finally, usability of a basic event, the rate of usability decrease in an object basic event by a measure and the implementation state of a measure are assigned as Table 2. Note that the value of usability of each basic event is 1 as a standard according to the definition in section 2.3.

The analysis based on the concept of phases in our method can calculate effective measures preferentially because a measure to recover usability of the violated phase

increases usability largely as compared with another one related to other phases. Therefore, our method can select appropriate measures consisting with actual service use.

The $U_{top}$, which is usability of the top event, gives

$$U_{top} = 1 - \prod_{c \in C} \left\{ 1 - \prod_{e \in Ec} U_e \prod_i \left( 1 - X_i \Delta U_{e,i} \right) \right\} \qquad (2)$$

where $U_e$ is the usability of a basic event $e$ and $\Delta U_{e,i}$ is the rate of usability decrease in $e$ by measure $i$.

## 4 RELATIONAL ANALYSIS WITH STATE TRANSITION DIAGRAM

We analyze relations among risks, usability and their chain relations with STD. FTA can also analyze these relations by combining FTs. However, the combining may make analysis complex and proper recognition of the relations difficult because FTs may become enormous and we may not identify where each event in FTs happens in an object network. In contrast, the method to make STD and incorporate with a network model can analyze risk, usability and these relations visually and intuitively.

### 4.1 Creating and Combining the STD of Risk and Usability

The STD of a risk as shown in Figure 4 is created by assuming a basic event in the FT at Figure 2 as an event in the STD and a result of the event of the FT as a state in the STD. Similarly, the STD of usability as Figure 5 is created from Figure 3.
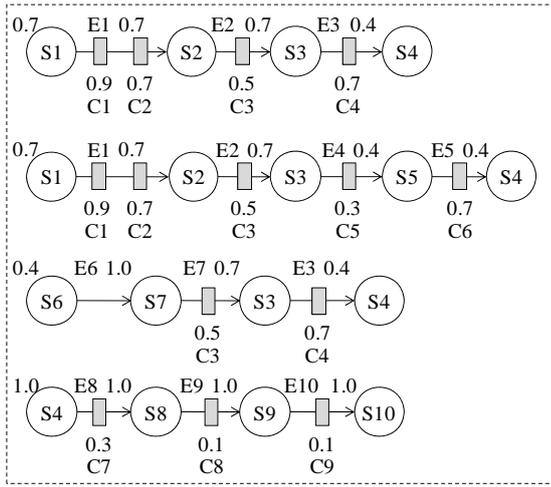
Figure 6: Combining of STDs.



Figure 7: Model of the network as shown in Figure 1.

Each arrow from one state to another has state transition probability, which is equal to probability/usability of a basic event as shown in Table 1 and Table 2, owing to the correspondence relation between FT and STD. In addition, measures to prevent from realizing a phase are placed on the arrow. Some measures can be placed on a single arrow.

The risk/usability decrease by a measure in Figure 4 and Figure 5 are also equal to the one in Table 1 and Table 2. The value is treated as the rate of blocking the state transition. Additionally, a state occurring independently such as motivation, "be willing to *do something*", has the occurrence probability.

The STDs can be combined by merger of the same state as shown in Figure 6. In order to combine the STDs, idempotent and distributive laws are applied because the STDs correlates with the FTs and states and events in the STD can be treated as constituent factors of risks and usability [9]. Note that commutative law cannot be applied because state transition has direction.

The STDs of both risks and usability are also combined by merger of the same states such as "PC A is usable" in both Figure 4 and Figure 5. In the case of, for example, leaking data from PC A by an e-mail, the value of 1 which is the standard of usability changes directly into success probability of the attack. Therefore, the STD of usability changes to that of risk.

The STDs are deployed on a network model. An organization has some network segments such as DMZ and intranet which may be divided into proper areas based on business departments. Based on the existing defense in depth model [11], our method creates a network model that has network segments and machines in each segment. The machines are also developed into some layers.

For example, a simple network configuration in Figure 1 changes into the model as shown in Figure 7 by separating
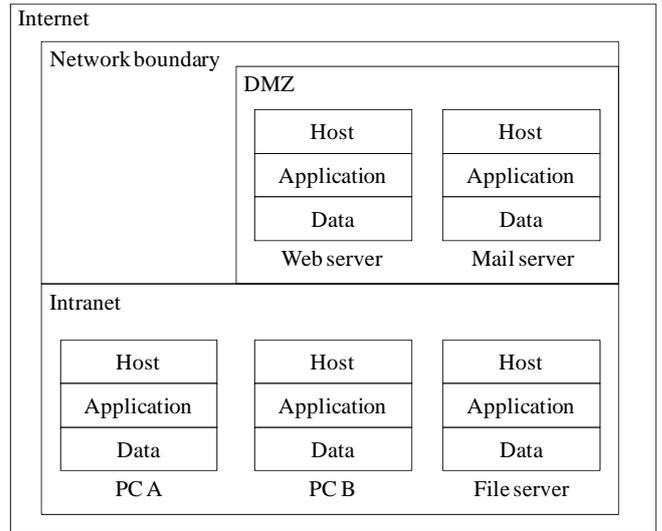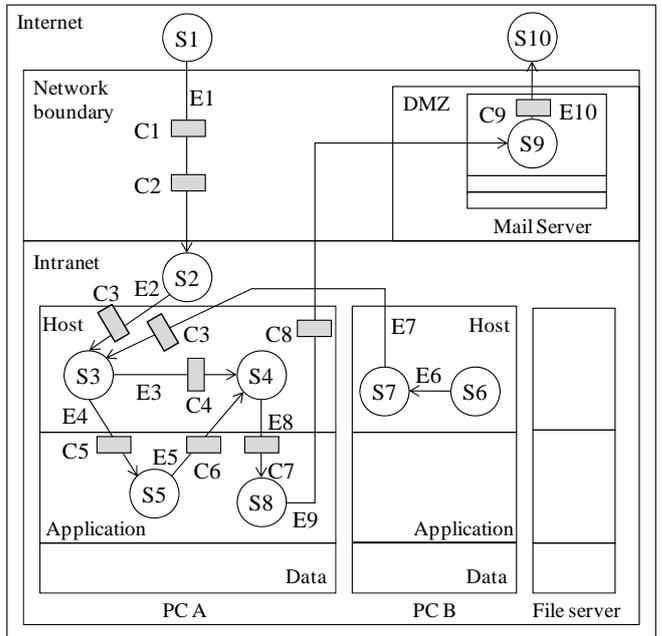


Figure 8: Deployment of risks, usability and measures on the network model.

the network into DMZ and intranet and developing each machine into the layers of host, application and data. The STDs are deployed on this network model as shown in Figure 8. Each state is on a layer to occur. Each arrow passes through some related layers. Each measure is set on a layer to implement.

In addition, STDs can be copied from one machine to others where machines in a common segment have same risks. Differences such as implementation states of measures among machines can also be customized as needed.

Details of the model should be decided depending on objects and accuracy of risk analysis that the organization requires. In the case of Figure 7, the firewall is excepted because we regard it as one of measures. The router is also omitted because we think it does not face threats. The

network model, for example, may treat several segments as a single area and add layers related to hardware such as memory devices like hard disks and I/O devices like LAN cards.

## 4.2 Risks and Usability Quantification

With Formula (1) and (2) by FTA, we quantify each risk and usability. On the other hand, STD expressing chain relations can also quantify them.

A risk probability, the value of a risk and usability can be quantified using a created model such as Figure 8. First, we decide an object risk and select states as starting points and ending point in order to quantify risk probability. At this time, several starting points can be selected. Next, we extract paths going from each starting point to the ending point. Then, the risk probability $P_n$ in path $n$ gives

$$P_n = \prod_{e \in E_n} P_e \prod_i \left(1 - \Delta P_{e,i} X_i\right) \qquad (3)$$

where $e$ is an event included in $n$ ($e$ include the occurrence of a starting point), $E_n$ is a set of $e$, $P_e$ is a state transition probability or the occurrence probability of $e$, $X_i \in \{0, 1\}$ is the implementation state of measure $i$ and $\Delta P_{e,i}$ is the rate of probability decrease in $e$ by $i$.

Finally, a total probability $P_{total}$, which means that at least one of the selected starting points arrive to the ending point, gives

$$P_{total} = 1 - \prod_{n \in N} \left(1 - P_n\right) \qquad (4)$$

where $N$ is a set of all paths.

In addition, Formula (3) and (4) equate to Formula (1) because STD has a correspondence relation to FT. Therefore, we must not exponentiate same state transition probability and the rate of risk/usability decrease by same measure in a case that several paths include a common event. We must replace exponent as follow [9]:

$$\left\{ P_e \prod_i \left(1 - \Delta P_{e,i} X_i\right) \right\}^2 \rightarrow \left\{ P_e \prod_i \left(1 - \Delta P_{e,i} X_i\right) \right\}$$

Furthermore, we can calculate the value of a risk in consideration of the value of assets. This paper define the value of risk as

*Value of a risk = Value of assets × Risk probability.*

For example, when a state on the data layer is selected as an ending point, state transitions arriving to the state are the risks about data. The value of risk is calculated using the value of data and risk probability calculated by Formula (3) and (4). When object assets are several data, we must calculate the value of the risk on a data-by-data basis because of differences of the value of data. In addition, when the implementation state and/or the rate of probability

Table 3: Transforming state transition probability to conditional probability.

| $P$(S1) |
| --- |
| $P_{S1}$ |

| $P$(S6) |
| --- |
| $P_{S6}$ |

| S1 | $P$(S2) |
| --- | --- |
| T | $P_{E1}$ |
| F | 0 |

| S6 | $P$(S7) |
| --- | --- |
| T | $P_{E6}$ |
| F | 0 |

| S2 | S7 | $P$(S3) |
| --- | --- | --- |
| T | T | $1-(1-P_{E2})(1-P_{E7})$ |
| T | F | $P_{E2}$ |
| F | T | $P_{E7}$ |
| F | F | 0 |

| S3 | $P$(S5) |
| --- | --- |
| T | $P_{E4}$ |
| F | 0 |

| S3 | S5 | $P$(S4) |
| --- | --- | --- |
| T | T | $1-(1-P_{E3})(1-P_{E5})$ |
| T | F | $P_{E3}$ |
| F | T | $P_{E5}$ |
| F | F | 0 |

\* $P$(S2), $P$(S3), $P$(S4), $P$(S5) and $P$(S7) are the conditional probability.
T means that the transition to the state is done.
F means that the one is not done.

decrease by a measure are (or should be) different from each data, we copy and customize STDs and calculate the risk probability and the value of the risk on an asset-by-asset.

Similarly, usability can be calculated by Formula (3) and (4) based on paths from starting points and an ending point. Note that $\Delta P_{e,i}$ is the rate of usability decrease in $e$ by measure $i$.

## 4.3 Causal Inference of Incidents

We can regard STDs such as Figure 8 as a model expressing cause-and-effect relations in a probabilistic viewpoint. Therefore, we can treat STDs as a Bayesian network with changing the state transition probability into the conditional probability when each state transition probability is independent of previous and following phases. As a result, we can infer the probability of causes of the occurrence of a risk event.

For example, in Figure 8, we suppose that the state S4 occurred and all measures had not been implemented. Table 3 shows the conditional probability of sates existing in the path to S4. The probability that S1 had occurred gives as follow:

$$P(S1|S4) = P(S1 \cap S4) / P(S4)$$

For another example, when we analyze closed object states such as {S2, S3, S4, S5}, there is no occurrence probability. In this case, we calculate the arrival probability of S2 as the occurrence probability with Formula (3) and (4).

In this way, our method can infer the probability of causes which are states existing on the turned paths from the target state.

## 5 EXPERIMENT

We experiment to apply our method at the stages of implementing and changing measures in a simple network example. We then confirm that our method can analyze phases of risks and usability and select optimal measures in consideration of their implementation place.

### 5.1 Assumption

In order to avoid complex analysis, we assume a simple network as shown in Figure 1. As mentioned at section 3.1, general network users browse web sites, send and receive e-mails and use files on a file server.

We analyze risks selectively because real results of risk analysis in actual organization networks are unavailable due to their confidentiality about security. Target risks are following: (i) unauthorized access to a PC via networks, (ii) data leakage of a confidential file via networks and (iii) virus infection on a PC. Target services relating to usability are following: (i) browsing web sites, (ii) sending e-mails, (iii) reading e-mails and (iv) using files in the file server.

In this experiment, each value is quantified in some levels. Regarding risks, the state transition probability of a phase has four levels (0.1, 0.4, 0.7, 1.0). The rate of risk decrease by a measure has five levels (0.1, 0.3, 0.5, 0.7, 0.9). Regarding usability, the state transition probability of a phase has the value of 1 as a standard. The rate of usability decrease by a measure has five levels (0.1, 0.3, 0.5, 0.7, 0.9). In addition, the implementation state of a measure has two levels (0, 1).

### 5.2 Risk Analysis and Measures Selection

**(1) The Model Creation**

First, we model the network. In this experiment, the network was modeled as shown in Figure 7.

Second, we analyze assets, threats, vulnerabilities and phases about target risks and usability like Figure 2 and Figure 3. Then, we clarify measures to basic evens based on [12] and assign the probability of a basic event, the rate of risk/usability decrease by a measure and the implementation state of a measure like Table 1 and Table 2. Finally, we express risks and usability as STDs and deploy them on the network model with measures. As a result, the model was created as shown in Figure 9.

In this case, all of the event name, the state transition probability and the risk/usability decrease by measures are eliminated from Figure 9 in order to avoid the complication of the model. Most parts of the STDs about the PC B are also eliminated because they were copied from the PC A.

**(2) Utilization of the Model for Risk Analysis and Measures Selection**

Some results of risk analysis and implement states of measures with FTA were reconsidered at the STD model creation.

First, several phases were modified to expand them in more detail because the model could not express several measures on an appropriate layer. The reason is that the phases of risks and the place to implement measures become clear. Specifically, we added the state named "inappropriate application use" and redeployed the measure named "application update" from the host layer to the application layer. In addition, we added the candidate of a measure named "limitation of usable applications" on the transition from "PC access" (on PC A) to "inappropriate application use". This measure, however, is not drawn in Figure 9 because we decided not to implement it.

Next, some of state transition probabilities were modified. Common phases in some risks had different probabilities each other even though the probability should be equal regardless of previous or following phases. The reason of these modifications is, similarly as above mentioned, that we can discover inconsistency of results of risk analysis by combining STDs.

Similarly, some of the rates of risk decrease by measures were modified. These values had been different from each measure even though the measures were clarified in common at analyzing risks. The reason of these modifications is that we can discover inconsistency of the rate of risk decrease by combining STDs.

Furthermore, some of the rates of risk decrease and implementation states of measures were added to insufficiently analyzed phases. We found that a measure effects to a specific target phase as well as to other phases. The reason of this work is that we can recognize whether a measure effects to other events passing through the same layer because the model can clarify phases of risks and the layer of a measure implementation. For example, we found that the measure named "access rights" for the risk of data leakage relates to control of a virus infection and sending/receiving attached files in an e-mail, and then added the measure on appropriate places of the data layer.

Finally, measures to effect some phases of risks in common are added as follows: "web filtering", "PFW preventing inbound access", "PFW preventing outbound access" and "limitation of attached files in e-mails". The reason of these additions is that we can consider whether there are measures effecting on some events. For example, the above "web filtering" effects on three phases efficiently.

**(3) Quantitative Assessment of Risks and Usability**

Table 4 shows the probability of risks and the value of usability at the initial analysis with only FTA and at the second analysis with both FTA and STD. We assessed the value of assets simply as shown in Table 5 and calculated the value of risks as shown in Table 6 from the probabilities and the values of the assets, which are amount of damage at the occurrence of risk events. Note that though these results include some probabilities/values of risks increase and usability decrease, they do not mean selected measures were inappropriate but our method could analyze risks and usability more accurately.

As a result, our method can assess risks and usability quantitatively and select measures to reduce risks and improve usability.
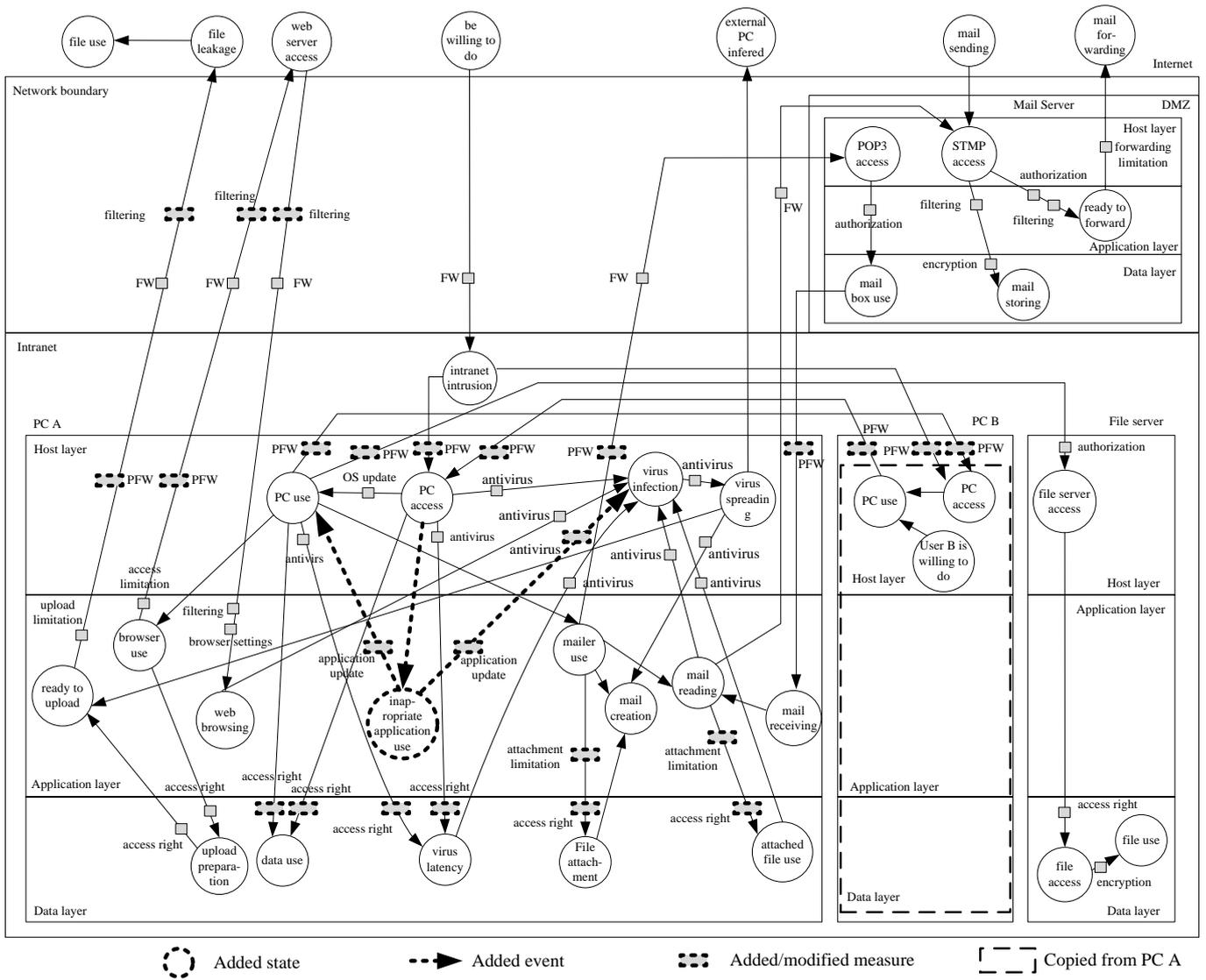
Figure 9: Modeling of risks, usability, and measures in the assumed network.

## 5.3 Analysis to Change Measures

For reviewing or changing measures, to identify shift of security and usability levels is important. In this situation, alternative measures may be needed.

Our method with FTA can calculate optimal measures to change measures by solving a discrete optimization problem with an objective functions to minimize risk increase or usability decrease and constraint functions to maintain levels of risks and usability. We can recognize the impact of changing measures visually by combining the above method with STD.

We suppose that, for example, a user of PC A requires opening a certain port of the firewall temporarily. The network becomes being intruded more easily by attackers via Internet if the firewall port is opened. At the same time, this change causes increasing probabilities to arrive all of the states continuing from the state named "intranet intrusion".

The additional measures selected with FTA in [7], although the results are not exact because of the difference

of an environment assumption, were as follows: (a) account lockout after login failure, (b) logging of PC access, (c) password protection on a screen saver and (d) prohibition of using HTML e-mail. Table 7 shows the risk probability, the value of risks and the usability at usual network operation, at firewall port opened and at above four measures added.

We can clarify that all of the selected additional measures are implemented on PC A using Figure 9. The port opening of firewall can be configured to effect only to PC A. Contrarily, a measure to the transition before or immediately following "intranet intrusion" should affect other users. On the other hand, the selected measures do not affect other users. Therefore, our method can select measures in consideration of their effects and the implementation layer.

In addition, another effective measure is "authentication for network access" (which is not drawn in Figure 9 because we decided not to implement it) deployed on the same transition as the firewall. The measure prevents intranet from being intruded. The other measure is "PFW" on the transition continuing directly from "intranet intrusion". The measure prevents the state from transiting to other states.

Table 4: Probability and usability at the first analysis with only FTA and reanalysis with both FTA and STD.

| Target of analysis | | FTA | FTA and STD |
|---|---|---|---|
| (R1) | Unauthorized access | 0.0285 | 0.0494 |
| (R2) | Information leakage | 0.0430 | 0.0198 |
| (R3) | Virus infection | 0.0181 | 0.0158 |
| (U1) | Browsing web sites | 0.810 | 0.590 |
| (U2) | Sending e-mails | 0.590 | 0.372 |
| (U3) | Reading e-mails | 0.590 | 0.413 |
| (U4) | Using files in the file server | 0.729 | 0.510 |

Table 5: Assumption of the value of assets.

(unit: yen)

| Use PC | 1,000,000 |
|---|---|
| Confidential files | |
| about product development | 50,000,000 |
| about customers | 10,000,000 |
| Not confidential files | 100,000 |

Table 6: The value of risks calculated from the value of assets and the probability.

(unit: yen)

| | Object | FTA | FTA and STD |
|---|---|---|---|
| (V1) | About user PC | 46,551 | 65,245 |
| (V2) | About all files | 2,581,968 | 1,188,176 |
| | total | 2,628,519 | 1,253,421 |

From a viewpoint of analysis, our method can identify potential states, which may occur by chain relations, by following STDs from a base state, which is directly caused by the transition that the changed measure had inhibited. Similarly, the method can also identify causes to raise the base state by tracing STDs back. At the same time, we can identify influences on usability.

From a viewpoint of selecting measures, in order to maintain (or reduce) risks, the method can narrow down the candidates of measures, depending on the concept of preventing chains or resolving causes of the risks. On the other hand, in order to maintain (or increase) usability, the method can also narrow them down, depending on the concept to improve usability of chain phases or resolve causes of usability decrease.

## 5.4 Probabilistic Causal Inference

We focus on the state named "PC A is usable", which means unauthorized use of PC A, as an example. First, paths reversed from the state are extracted from the STDs in Figure 9. The result is shown in Figure 10. Note that Figure 10 includes measures which are not implemented and limits the states on the PC B to S6, S7 and related events. In addition, we connect an event directly from S2 to S7 in order to make the STD simpler though we should analyze

Table 7: Shift of risks and usability caused by changing measures.

probability and usability

| | At usual operation | At the FW port opened | At other measures added |
|---|---|---|---|
| (R1) | 0.0494 | 0.0796 | 0.0518 |
| (R2) | 0.0198 | 0.0375 | 0.0269 |
| (R3) | 0.0158 | 0.0215 | 0.0195 |
| (U1) | 0.590 | 0.590 | 0.590 |
| (U2) | 0.372 | 0.372 | 0.335 |
| (U3) | 0.413 | 0.413 | 0.372 |
| (U4) | 0.510 | 0.510 | 0.510 |

value of risk

| (V1) | 65,245 | 101,094 | 71,298 |
|---|---|---|---|
| (V2) | 1,188,176 | 2,251,166 | 1,619,260 |
| total | 1,253,421 | 2,352,261 | 1,690,558 |
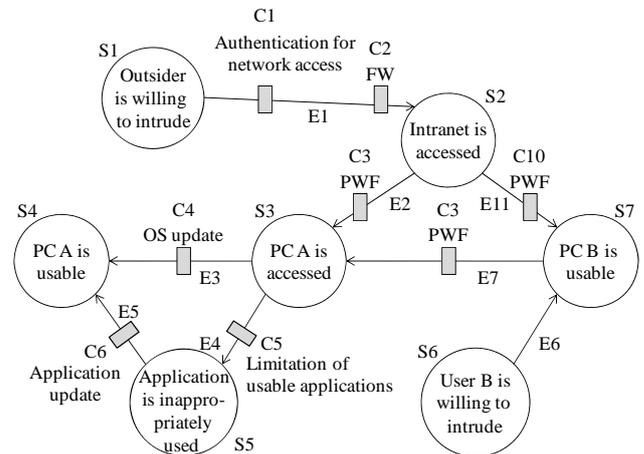


Figure 10: Candidates of the causes in unauthorized use of PC A.

events of accessing to and intruding into the PC B via Internet.

Next, we change state transition probability to conditional probability and infer causes of the incident. The state transition/occurrence probabilities, the rate of risk decrease and the implementation states of measures were analyzed at section 5.2 as shown in Table 8.

Then, we infer causes of S4. Table 9 shows the probability of each state which had occurred before S4 occurred at usual network operation, at a certain port of firewall (C2) opened and at "authentication for network access" (C1) added respectively.

At usual network operation, because the firewall protects from inappropriate access via Internet, outsiders might be willing to intrude into intranet (S1) but the probability of successful intrusion was inhibited (S2). On the other hand, the probability that the PC B was used (S7) and the user B was willing to intrude (S6) were high. These probabilities mean that an internal user is more suspicious.

Table 8: Assigned values of states, events and measures at the analysis.

| state or event | state transition /occurrence probability | measure | risk decrease | state |
|---|---|---|---|---|
| S1 | 0.7 | - | - | - |
| S6 | 0.4 | - | - | - |
| E1 | 0.7 | C1 | 0.9 | 0 |
|  |  | C2 | 0.7 | 1 |
| E2 | 0.7 | C3 | 0.5 | 1 |
| E3 | 0.4 | C4 | 0.7 | 1 |
| E4 | 0.4 | C5 | 0.3 | 0 |
| E5 | 0.4 | C6 | 0.7 | 0 |
| E6 | 1.0 | nothing | - | - |
| E7 | 0.7 | C3 | 0.5 | 1 |
| E11 | 0.7 | C10 | 0.5 | 1 |

At a certain port of the firewall opened, outsiders can access intranet more easily via Internet. The probability of intrusion by outsiders (S2) became higher drastically. This result means that an outsider is more suspicious.

At "authentication for network access" added, the probability of intrusion by outsiders (S2) decreased significantly. The probability that the PC B was used (S7) and the user B was willing to intrude (S6) became extremely higher. These probabilities mean that an internal user is highly suspicious.

Note that the attacker had surely accessed to the PC A (S3) before he/she operated it without authorization (S4). On the other hand, intruding into the PC A (S5) is absolutely not related to whether vulnerabilities of the operating system or of an application is exploited, which means whether the transition from S3 to S4 goes through S5 or not. Therefore, the following conditional probabilities are constant.

$$P(S3|S4) = 1, \quad P(S5|S4) = 0.72$$

As a result, our method can infer probabilistic causes of risks. Therefore, we can use the probabilities as one of information to select network monitoring points for proactive measures and to analyze causes of incidents for reactive ones.

# 6  EVALUATIONS

We evaluate our method for resolution of research issue described in chapter 2.

## 6.1 Analysis of Risk and Usability with Chain Relations

In the experiment, we analyzed phases of risks and services use with FTA and converted to STD. we clarified the chain relations among risks by combining the STDs. We also analyzed the phases of services use and clarified

Table 9:  Probability of causation of unauthorized use of PC A.

|  | At usual operation | At the FW port opened | At authentication added |
|---|---|---|---|
| $P(S1|S4)$ | 0.78 | 0.86 | 0.73 |
| $P(S2|S4)$ | 0.38 | 0.77 | 0.15 |
| $P(S3|S4)$ | 1.00 | 1.00 | 1.00 |
| $P(S5|S4)$ | 0.72 | 0.72 | 0.72 |
| $P(S6|S4)$ | 0.8 | 0.59 | 0.92 |
| $P(S7|S4)$ | 0.9 | 0.78 | 0.96 |

relations among usability.

In addition, the STDs of risks and usability were combined. The combined STDs clarify the relation among risks and usability. Moreover, the inconsistencies among risk analysis results are discovered and modified with combined STDs.

As a result, we confirm that our method can analyze risks and usability with their phases in consideration of chain relations. Furthermore, our method can assess risks and usability by calculating the risk probability, the value of risks and the usability as the result of section 5.2.

## 6.2 Analysis of Relations among Risks, Usability and Measures

In the experiment, the combined STDs were deployed on the network configuration model based on defense in depth. The incorporated model could express correspondence relations among phases of risks and usability.

In addition, measures clarified with FTA are developed on the appropriate layer in the incorporated model. We can recognize clearly the implementation place of measures.

On the other hand, our method can analyze relations among risks, usability and measures because influence on risks and usability at changing measures became clear in section 5.3.

## 6.3 Selecting Optimal Measures and Causal Inference

At first measures selection in the experiment, we could seek efficient measures preventing some of objective risks by analyzing the risks based on phases. We also visually found inconsistencies among the results of analysis and lack of requisite measures. As a result, the method can analyze effectiveness and the implementation place of measures exactly and select optimal measures.

At changing measures, we recognized the influence on risks and usability visually and quantitatively. We can select measures based on the concept of preventing causes or chains, which mean prior or latter phases. Consequently, our model can select optimal measures to inhibit risks increase and/or usability decrease.

On the other hand, we inferred probabilistic causes of a risk. We can detect incidents by monitoring events with high probability of causes and/or machines including some of the

events. In addition, we can identify causes of incidents efficiently by investigating such events mainly.

# 7   CONCLUSION AND FUTURE WORKS

We have proposed the method to analyze risks and usability in consideration of chain relations and select optimal measures. The method analyzes each risk and usability with FTA, converts the results to STDs and deploys the STDs on the network model based on defense in depth. The method can be converted to a Bayesian network and thereby it can infer causes of incidents.

Through the experiment, we confirm that our method can quantitatively and intuitively (i) analyze risks and usability, (ii) select optimal measures and monitoring points and (iii) be available to trace causes of incidents.

The occurrence probability of a state, the state transition probability and the rate of risk/usability decrease by measures must be exact to assess risks and usability correctly. However, to assign proper value is difficult because these values may be different from each environment or each user. One of the solutions to solve this problem is to cycle risk analysis, assessment and review through network operation. The other solution is that stakeholders decide the values through negotiations, generally called risk communication such as [4].

Moreover, Figure 9 is too complicated even though we treat only three risks and four services. In the case of analyzing more risk and usability, the number of states and transitions may become huge. In contrast, the increase of the number of states may converge at certain number because risks/usability related to the same layer often have a common transition. Additionally, each state transition must be independent of the previous and next transitions in order to infer probabilistic causes. One way to achieve this state independency is to involve some parameters in each state and each event, such as time of transition and a person who transits the states. However, excessive number of parameters may increase the number of states and transitions. We will study more about this problem.

## REFERENCES

[1] ISO/IEC 27002. http://www.iso.org/iso/home.htm
[2] A. Zuccato, A Decision Matrix Approach –to Prioritize Holistic Security Requirements in E-commerce, Security and Privacy in the Age of Ubiquitous Computing: IFIP TC11 20th International Information Security Conference, pp.35-49, Springer (2005).
[3] B.-C. Guan et al., Evaluation of Information Security Related Risks of an Organization –the Application of the Multi-criteria Decision-making Method, Proceedings of IEEE 37th Annual International Carnahan Conference on Security Technology, pp.168-175 (2003).
[4] H. Yajima et al., Evaluation of the Participant-Support Method for Information Acquisition in the "Multiplex Risk Communicator", LNCS, Vol.4558, pp.195-203 (2007).
[5] I. Kotenko and M. Stepashkin, Attack Graph Based Evaluation of Network Security, LNCS, Vol.4332, pp.216-227 (2006).
[6] L. Wang et al., Measuring the Overall Security of Network Configurations Using Attack Graphs, LNCS, Vol.4602, pp.98-112 (2007).
[7] K. Kato and Y. Teshigawara, A Proposal of Selecting Optimal Countermeasures with Security and Usability in a Special Network Use, IPSJ Journal, Vol.49, No.9, pp.3209-3222 (2008) (in Japanese).
[8] K. Kato and Y. Teshigawara, A Study on a Model of Risks and Countermeasures Based on the Concept of Defense in Depth, Multimedia, Distributed, Cooperative, and Mobile (DICOMO) Symposium 2008, pp.1531-1540 (2008) (in Japanese).
[9] J. D. Andrews and T. R. Moss, Reliability and Risk Assessment -Second Edition, Professional Engineering Publishing (2002).
[10]   Microsoft TechNet, Security Content Overview. http://technet.microsoft.com/en-us/library/cc767969.aspx
[11]   Microsoft TechNet, Chapter 3: Antivirus Defense-in-Depth.   http://technet.microsoft.com/en-us/library/cc162798.aspx
[12]   ISO/IEC 13335. http://www.iso.org/iso/home.htm

# Invited Speech: Dr. Norman Abramson
## (Chair Yoshimi Teshigawara)

# IWIN 2009

### History of the AlohaNet - ALOHA to the Web

Honolulu, September 15, 2009

**Norman Abramson**
Professor Emeritus -- University of Hawaii
norm @ hawaii.edu

---

# SUMMARY

- **The history of ALOHA**

- **ALOHA applications**

- **The ALOHA channel**

- **Current issues in ALOHA**

# THE HISTORY OF ALOHA

Honolulu

Hawaii

1968

# THE ALOHA SYSTEM

ATS-1

INTELSAT IV

packet
repeaters

packet
sensors

terminals

PCU
PCU
PCU
PCU

MODEM

SIMP

IBM 370/158

HP
2100

T
R

PCU

M
P
L
X

multiplexors

PCU

M
P
L
X

MINI

minis, RJE
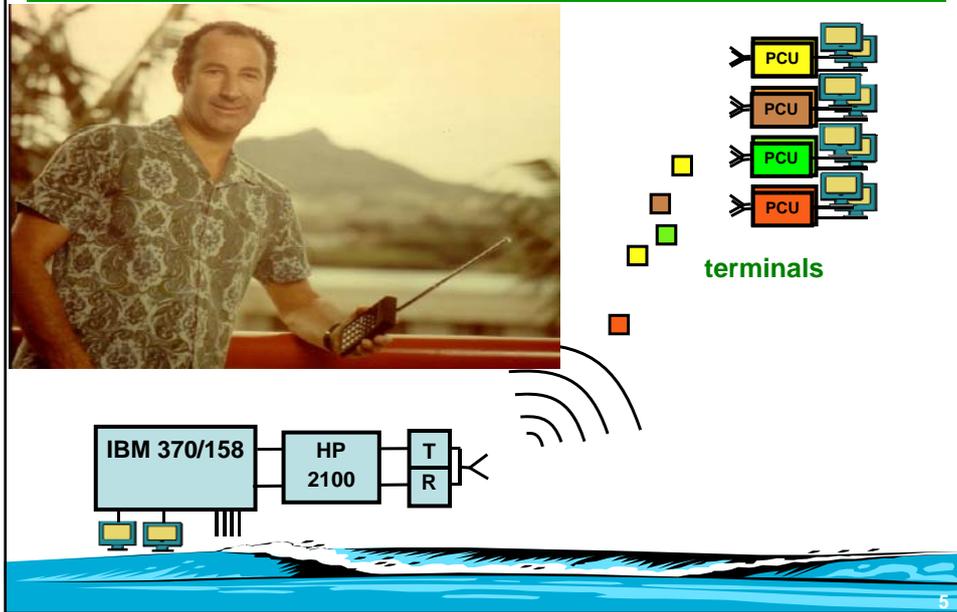
# THE ALOHANET



# SUMMARY

- **The history of ALOHA**

- **ALOHA applications**

- **The ALOHA channel**

- **Current issues in ALOHA**

# ALOHA APPLICATIONS

1. wired and fixed wireless (LANs)

2. satellite networks

3. wireless mobile networks

---

# LANs – ETHERNET

- Alto ALOHA Network – Metcalfe '73

- CSMA/CD – Kleinrock / Tobagi '75

- Standard – IEEE 802.3 '83 ......
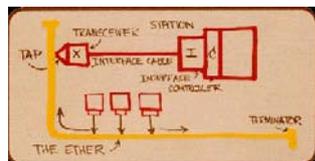


**1973**



**2009**

# LANs – WiFi

- **unlicensed bands (US) '85**

- **CSMA/CA – Kleinrock / Tobagi '75**

- **standard – IEEE 802.11 '97 ......**

**1995**



**EtherNet or ALOHANet ?**

---

# LANs – WiFi  (Applications)

# SATELLITE ALOHA

**Marisat
1976**
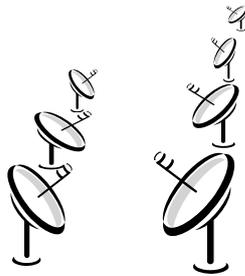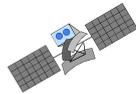
**private networks
1980's**

**Internet access
2000's**

# ALOHA in MOBILE NETWORKS

**1980's
1G**

**1990's
2G**

**2000's
3G**

**2010's
4G ...**

request
control

*

SMS

*

GPRS

*

web

# SUMMARY

- **The history of ALOHA**

- **ALOHA applications**

- **The ALOHA channel**

- **Current issues in ALOHA**

# THE ALOHA CHANNEL



packet

collision

repetition

time
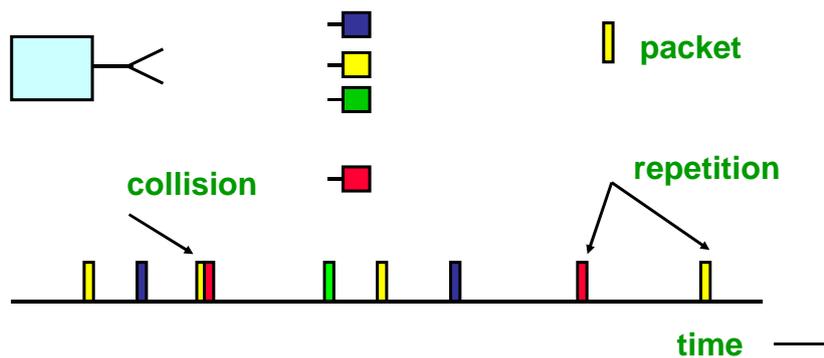
$$\text{maximum throughput} = \frac{1}{2\,e} = 0.186$$

# ALOHA  FLAVORS

- unslotted ALOHA

- slotted ALOHA

- reservation ALOHA

- CSMA

- spread ALOHA

15

# SUMMARY

- The history of  ALOHA

- ALOHA  applications

- The  ALOHA  channel

- Current issues in  ALOHA

16

# CURRENT ISSUES in ALOHA

- **to slot or not to slot ?**

- **spread ALOHA – CDMA w/o CD**

- **WiFi – EtherNet or ALOHANet ?**

- **bursty traffic – charging by the byte**

- **throughput and protocol overhead**

17

# REFERENCES

● N. Abramson, *The ALOHA System - Another Alternative for Computer Communications,* Proceedings of the 1970 Fall Joint Computer Conference, AFIPS Press, Vol. 37, 1970, pp. 281-285.

● L. G. Roberts, *ALOHA Packet System With and Without Slots and Capture*, Computer Communications. Review, Vol. 5, No. 2, pp. 28-42, Apr. 1975.

● R. M. Metcalfe, *Packet Communication*, Massachusetts Institute of Technology, Cambridge, MA, Report MAC TR-114, July 1973.

● N. Abramson, *The Throughput of Packet Broadcasting Channels*, IEEE Transactions on Communications, Vol. Com-25, No. 1. Jan. 1977, pp 117-128.

● R. Binder, et al., *ALOHA Packet Broadcasting – A Retrospect*, AFIPS Conference Proceedings, 1975 National Computer Conference, AFIPS Press, Montvale, New Jersey, May 19-22, 1975.

● N. Abramson, *VSAT Data Networks*, Proceedings of the IEEE, Special Issue on Satellite Communications, Vol. 78, No. 7, pp. 1267-1274, July, 1990.

● N. Abramson, *The ALOHANet – Surfing for Wireless Data*, IEEE Communications Magazine, to be published 2009.

18

# HISTORICAL REFERENCES

● N. Abramson, *The ALOHA System - Another Alternative for Computer Communications,* Proceedings of the 1970 Fall Joint Computer Conference, AFIPS Press, Vol. 37, 1970, pp. 281-285

● R. Metcalfe, *Xerox PARC memo*, from Bob Metcalfe to Alto Aloha Distribution on Ether Acquisition, May 22, 1973.

● N. Abramson, *The ALOHA System Final Technical Report*, Advanced Research Projects Agency, Contract Number NAS2-6700, October 11, 1974

● R. Binder, et al., *ALOHA Packet Broadcasting – A Retrospect*, AFIPSConference Proceedings, 1975 National Computer Conference, AFIPS Press, Montvale, New Jersey, May 19-22, 1975.

● N. Abramson, *The ALOHANet – Surfing for Wireless Data*, IEEE Communications Magazine, to be published 2009.