IOWA STATE UNIVERSITY
Digital Repository

2015

An empirical study of communication media and issues in globally distributed software development

Ya Chen
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd

Part of the Computer Sciences Commons

An empirical study of communication media and issues in globally distributed software development

by

Ya Chen

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Computer Science

Program of Study Committee:
David Weiss, Major Professor
Carl K. Chang
Les Miller

Iowa State University

Ames, Iowa

2015

TABLE OF CONTENTS

ABSTRACT

It is known that distributed software development is risker and more challenging than collocated software development. Previous research has suggested that there are several factors that cause delay in distributed development including cultural differences, lacking common ground, communication and coordination overhead [4]. We use data collection forms and interviews to gather data from a course on distributed software development (DSD), collaborated among four universities in four widely-separated countries to do an empirical case study. We try to evaluate communication media effects on DSD and verify hypotheses and issues about communication among distributed sites. We found that selecting correct communication media is helpful in solving communication issues and decreasing development delay. We also evaluate the effects of common ground, human resources, team culture, module structure distribution and time zone on project time and effort. Further, we found several communication issues in this DSD case study and present a strategy to improve DSD and avoid these communication issues.

CHAPTER I

INTRODUCTION

A distributed software development (DSD) is a software project done in a multi-site, multicultural, globally distributed environment. Project members may not see each other face to face but they are all working collaboratively toward the outcome of the project. Often the communication in DSD is done through email, IM and internet-based collaborated development tools [1].

1.1 Why choose DSD

Globally distributed software development is an increasingly common strategic response to issues such as skill set availability, government restrictions, acquisitions, increased code size, cost and complexity and other resource restrictions [2, 3]. Nowadays, DSD is increasingly used by companies of all but the smallest size. Over half of the fortune 500 companies use DSD as their main project development strategy.

Why do these companies choose DSD? There is now considerable foreign talent available that can help companies expand their pool of trained workers. In addition, DSD is a way to get closer to customers and use locality specific expertise to customize or localize products. If government is a customer requiring companies' R&D help in a country, companies may get some favorable tax treatment based on national policy. Besides the lower financial cost, round-the-clock development in

DSD could lead to shorter intervals, which decreases time cost for companies. Also, some countries may require a company to have a local office.

1.2 Communication in DSD

DSD is still faced with many challenges not inherent in collocated teams such as delayed feedback, restricted communication, lacking common ground, less shared project awareness, difficulty of synchronous communication, inconsistent development and build environments lacking of trust and confidence between sites [4]. There were some studies that examined these delay factors associated with DSD and direct reasons for them. Herbsleb and Mockus reported differences between same-site and distributed social networks, testing several hypotheses about characteristics of distributed social networks that may be related to delay [5]. There was also an empirical study of windows Vista along with post-release failure information to evaluate the hypothesis that globally distributed software development leads to more failures. As a result, they found a negligible difference in failures between DSD and development by collocated teams. [6]

What we focus on are communication issues in DSD. In fact, software engineers at cross-sites spend a large proportion of their time on communication [16]. Thus, communication is a challenge even in collocated software development environment and becomes even more problematic for DSD projects. Challenges in communication slow down the overall project process. In an empirical study of time use of developers in a large software engineering organization, Perry et al [7] found

that "one of the most salient impressions conveyed by observation was the sheer amount of time each developer spent in informal communication". In the study, the developers spent an average of 75 minutes each day in "unplanned interpersonal interaction". Developers spent considerable time on unnecessary communication, perhaps too much time. Previous research also suggests that cross-site communication issues cause a substantial loss of development speed. In an empirical study of speed and communication in DSD, Herbsleb and Mockus [5] reported that "we investigate relationships among delay, communication, coordination and geographic distribution of work" T.J. Allen also observed that once there's about 30 meters of distance in between employees, collaboration drops completely [17].

Beside the communication issues, we also focus on evaluating the communication media in DSD. In this study, the communication media we focus on are emails, Instant Message (IM), face-to-face communication, phone, video communication tools and development collaboration platforms such as Assembla. Here face-to-face communication means the communication is done in the same place, which means people need to get together for communication. Previous research proposed a media richness theory [8], which is based on the concept of richness of a medium: "richer media should be used for tasks of higher uncertainty and complexity, while simple and uncomplicated tasks should be dealt with leaner communication media [sic]". We also evaluated this point in our case study. On communication media, Tuomas et al [9] analyzed communication tools usage and found that "When working in a distributed setting, it is important to make sure the communication tools are

compatible with the project and the team. Both technically and socially." They also found that "It is important to have a comprehensive selection of communication tools available to cover the communication and collaboration needs of a DSD team."

1.3 Research Questions

Based on the previous research we mentioned above, communication is a major risk in DSD. What we are focusing on in this thesis is communication in DSD and how to improve communication efficiency. This thesis reports a study of evaluating communication tools and issues through a DSD course project during two semesters. This DSD course is held at Iowa State University in the fall semester each year. We evaluated and investigated some strategies and solutions in order to shed light on possible causes of communication issues and delay.

Our Research Questions are:
1. What strategy can improve communication efficiency in DSD projects?
2. How can communication media be efficient in DSD projects? Are there any differences in the way people use communication media between DSD and co-located development?
3. What communication issues do DSD projects have? How to mitigate or solve these issues?
4. Does DSD have the same communication issues as co-located development?

CHAPTER II

STUDY CASE

2.1 Introduction to Study Case

In this chapter, we introduce the data set used in this study and provide the justification for using this data set. The data set was originally collected from a one-semester course on distributed software development, taught as a collaboration among four universities in four widely-separated countries (COMS510 - Distributed Software Development at ISU). The course is offered independently by each university, and their students collaborate to produce a software application. Students were expected to develop a software application by cooperating with students at the other universities, surmounting differences in geographic location, time zone, culture and native language. The data were collected across two different offerings of the course during a two-year (Fall 2013, Fall 2014) interval, and were collected under close scrutiny. We monitored and validated the data weekly. The data serve to verify hypotheses proposed in this study.

2.1.2 The data set in Fall 2013

For the Fall 2013 data set, 31 students in total took COMS510 from four universities in different countries, including Iowa State University (ISU - Ames, United States), Ji Lin University (JLU – Chang Chun, China), National University of Colombia (UNAL - Bogota, Colombia) and King Mongkut's University of

Technology Thonburi (KMUTT - Bangkok, Thailand). Students are organized into teams, each team local to one university, and each team responsible for one module in the project. Team composition is shown in Table 1. Each team has six roles, which are project manager, liaison (usually assigned to the project manager), system engineer, architect, software engineer, and tester.

Table 1. Team composition in Data Set

	Fall 2014		Fall 2013	
	Students	Teams	Students	Teams
ISU	18	3	10	2
JLU	12	2	3	1
UNAL	8	2	6	1
KMUTT	12	2	13	2

The project they developed in Fall 2013 is a class attendance system, whose purpose was to help instructors to record students' attendance automatically through face recognition. Instructors can use a client application (Web client) to request that a classroom camera take pictures for the purpose of recording students' attendance. The client will identify students in the classroom through the use of face recognition technology; it then records students' attendance automatically into an attendance database that can be queried. A client application could use the camera to identify students in a class picture and display their names in real-time for the instructor on his/her laptop. The desired result is that the instructor is able to recognize students by name during the class and that the instructor and students could have available a record of who attended which classes.

The module structure is shown in Figure 1. Each team is in charge of one module in the project. The client module provide UI to operate the system by instructors and students. Here the client is only web based. The facial recognition server module serves to recognize the pictures provided by clients and returns recognition results to the facilitator module. Only the pictures of people whose pictures have been trained in the server module can be recognized. The Facilitator module is in charge of coordinating the services between clients and servers. The Facilitator module tests servers to know which server is available and compares the similarity and quality of results from multiple servers. The attendance database module is in charge of recording the attendance of students at class.
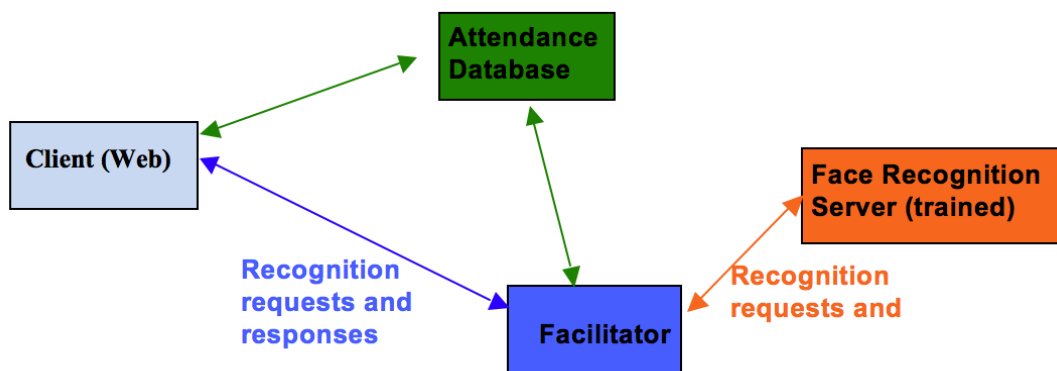
Figure 1. Module structure of class face recognition attendance system

2.1.2 The data set for Fall 2014

This data set includes 50 students who participated in the COMS510 course at four universities (ISU, JLU, UNAL, KMUTT). Team composition is shown in Table 1.

Students in JLU and UNAL are undergraduate students with less software development experience, while students in ISU and KMUTT are graduate students with more experience. In addition, most of the students at KMUTT have 2 or 3 years industrial working experience.

The project at Fall 2014 is also a face mapper system, whose purpose was to record the attendance of students on classes or the persons who attend a meeting. The main function is similar to the system developed in Fall 2013. However, students developed the system from scratch, and they did not have previous development as references. Additional functionality is a face map that shows the position of persons on the picture. Besides the web clients, instructors and students also can check records and operate the system on mobile clients.

The project's module structure is shown in Figure 2. Each team is in charge of one module. Similarly, this project also includes client module, facilitator module, attendance database module and server module. However, besides the web based client, there is also a client for smart phones. Compared with the module structure for Fall 2013, we added a face map module in this system. The face map module is able to display the positions of the persons in the pictures.
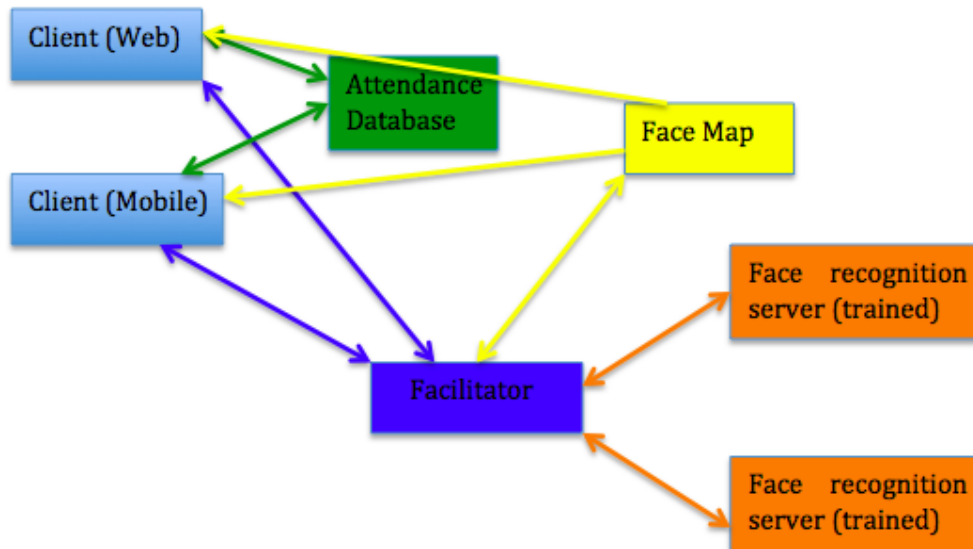
Figure 2. Module structure of Face Mapper project

2.1.3 Why did we choose this data set?

In order to explore our research on communication in DSD, we chose this course as our study case. On the one hand, the course is mimicking an industrial software development environment. On the other hand, we can collect the data easily and validate it in time. In order to help students finish the project, we constructed a Goal, Questions, Matric (GQM) graph, which is shown in Appendix B. GQM is an efficient approach to software metrics, which defines a measurement model on three levels: conceptual level, operational level, quantitative level [16]. The data collection is an accurate and validated evaluation way to show performance of the project.

The study case is a distributed software development (DSD) project, with characteristics typical of such projects. The four universities involved are located at four different countries with different cultures and native languages. The biggest time

zone difference between two sites is 14 hours. The detail of sites distribution is shown in Figure 3.   Although the students have less software development experience than many industrial software developers, the course project can emulate industrial DSD in some ways. The data obtained from this project reflects a similar DSD situation in industry and helps us to explore and solve associated research questions.
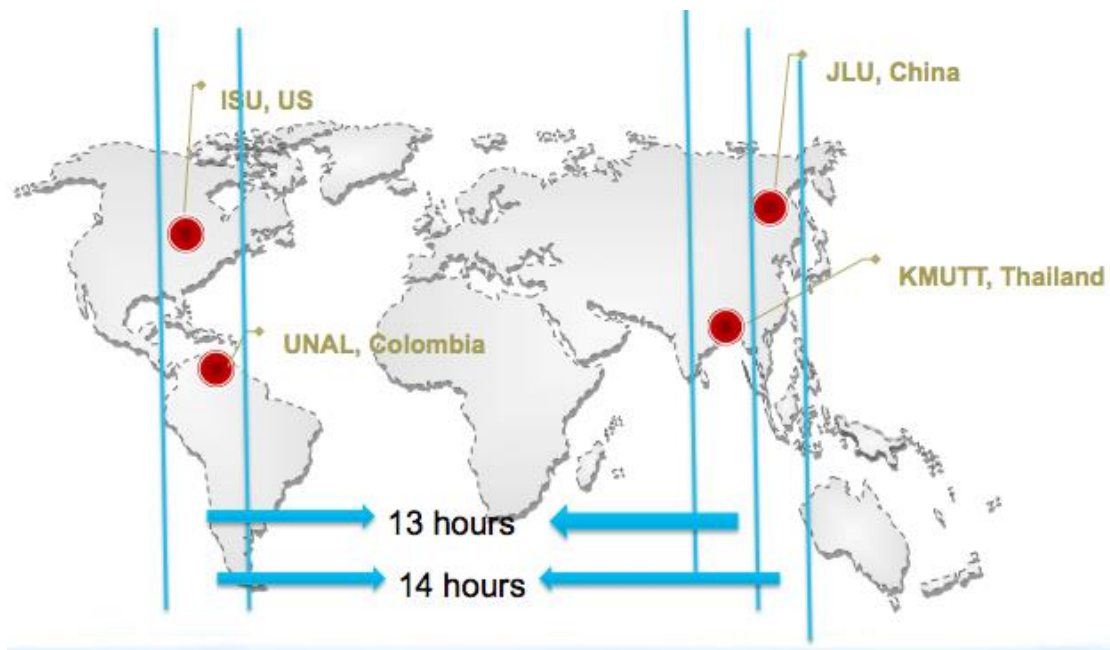
Figure 3.   The distribution of four universities

The data are easy to get and validate. We can collect the data easily and validate the data weekly because data collection gets supports from all the instructors in all of the universities. The students tend to follow the instructors' instructions, so we require students to update their data collection weekly. Meanwhile, once we find any data that seems abnormal or strange, we contacted the student who provided this data to make sure the data are correct or to determine what abnormal situation occurred.

2.1.4 What are the differences between these two data sets?

Compared with the data in Fall 2013, the data in Fall 2014 are more complete and accurate. Because Fall 2013 was the first time we collected the data, we did not consider validation as a serious issue and did not validate data in time. Some data were not updated weekly and some data are missing. Also, the data from Fall 2013 focused on collecting information on distributed communication media and did not collect data about local interactions and issues. In order to make sure the data are complete, we asked some students to recall the missing data. However, the missing data constituted only 11% of all data and most of the missing date can be recalled by students who filled out the data, so we do not think this issue is a big problem for accuracy. Because of our experience in Fall 2013, the data collection from Fall 2014 is better. Besides recording the information on cross-site and local site communication media, we also recorded the details of communication issues in the project, including what the communication issue was, how to solve it, what communication media caused the issues and what media are helpful to solve the issues. We improved both sheets used to collect the data. We used Google docs to collect the data in Fall 2014 in order to improve the efficiency of data collection, and so that all issues in one team could be tracked on one sheet.

2.1.5 Validating the accuracy of the data

To ensure the data are updated in time, we collected the data weekly and monitored the data weekly in the process. Once we found any strange data or the data

showed an abnormal situation, we would contact the students who filled out the data and figure out what caused the abnormal data. All the instructors in the program also held a weekly meeting to discuss the progress of the project including validating the data. We also asked instructors to give a detailed explanation for the abnormal data. Its common for different sites to have different available time slots, which affects the accuracy of the data. For any event issues, we have recorded these situations in our data collection.

2.2 Methods

In this section, we describe our methods of gathering data for our case study. In order to verify our hypotheses correctly and precisely, in-person interviews is also a method we used to collect data.

2.2.1 Data collection

In order to gather data, we designed a data collection form and asked students to fill out the form weekly. Each team has one data collection sheet, so their PMs are in charge of monitoring the data collection completion. Every person in the team needed to fill out this form, and was not allowed to write on others' behalf. For the Fall 2013 data set, the data collection form was collected and merged by Project Managers (PM) of each team through email weekly. At Fall 2014, we decided to collect data on Google docs, so the students could update their data in time and conveniently. The data collection sheets are mainly to collect the time effort and issues.

Time Effort

      For the time effort part (see Appendix A), the form records the weekly time effort in hours that each person spent on the project. This form also records the role of the person, so we can see if there is an association between project effort and team roles. The most important part of this data collection is gathering the data for media tools. Our data collection form records the time effort (by hours) on media tools usage for both local site and cross-site communication. Local site media includes email, face to face meeting, telephone, Instant Message (IM) and platform (See Appendix A). Here platform means some software development collaboration platform such as Assembla (http://www.assembla.com/), OKR (Inner software development management tool in Google) or Asana (https://app.asana.com/). In the project, we used Assembla as our collaboration platform between teams. Face-to-face meeting means all participants are at the same place and students are communicating face to face. When students used these media for communication at the same place, the time effort was recorded

      On cross-site data, the media include Emails, face-to-face meeting, IM and platform. We do not choose telephone as media because students are not able to afford the cost of international phone call.

Issues

      For the issues sheet (See Appendix A), all communication issues are recorded. The data includes the description of each issue, the time spent on this issue, which communication media caused this issue or is helpful to solve this issue, and the

description of the solution. Each team only has one issue sheet, so all team members in the team are able to track if the issue they currently encountered already existed. If the detail for this issue is not complete, team members can update the issue with more information. Every student also needs to record any time they spent on each issue in their time effort sheet. We also can track which specific issue team members are involved at the same time.

Why did we gather this data weekly? Students are required to update their data weekly. If they are recording their time effort or other data later, it's hard for them to recall the precise data. In addition, we also can monitor the data weekly. Once we found any unusual data, we asked students for specific reasons. Every week all instructors participated in a Skype meeting weekly. If there were any issues or problems in the process, they were discussed in detail at this weekly meeting and solutions were proposed.

2.2.2 Interview

Although the data collection sheet is the main method for collecting data, we also used in-person interviews to help verify hypotheses. At ISU, we have face to face interviews. For other sites, we set up Skype meetings with the students. The interview usually lasts 20~30 minutes. Overall, 19 students participated in interviews during this study. We tried to cover all teams in this program and talked with at least one student in each team. Table 2 shows the number of interview participants by location during the semester.

Table 2. Number of interview participants by location.

Site	Fall 2013	Fall 2014
US	4	4
China	1	2
Columbia	2	2
Thailand	2	2

CHAPTER III

ANALYSIS AND RESULTS

In this chapter, we describe the data analysis and its results, and verify the hypotheses based on the research questions. Then we summarize the communication issues during this case study and how to solve and mitigate them.

3.1 Verification of the hypotheses

Hypothesis 1. In DSD, teams tend to communicate through text based media. In co-located development, teams tend to communicate through audio media.

In order to figure out how to use communication media in co-located development and DSD, we chose different media as study targets. In DSD, we focus on the time effort of Email (cross-site), IM chat (cross-site), Video meeting and platform. In co-located development, we monitor the time effort of Email (local site), IM chat (local site), phone and platform. Since team members can use platforms on cross-site and local site, we considered platforms as one media in the analysis of DSD and collocated development.
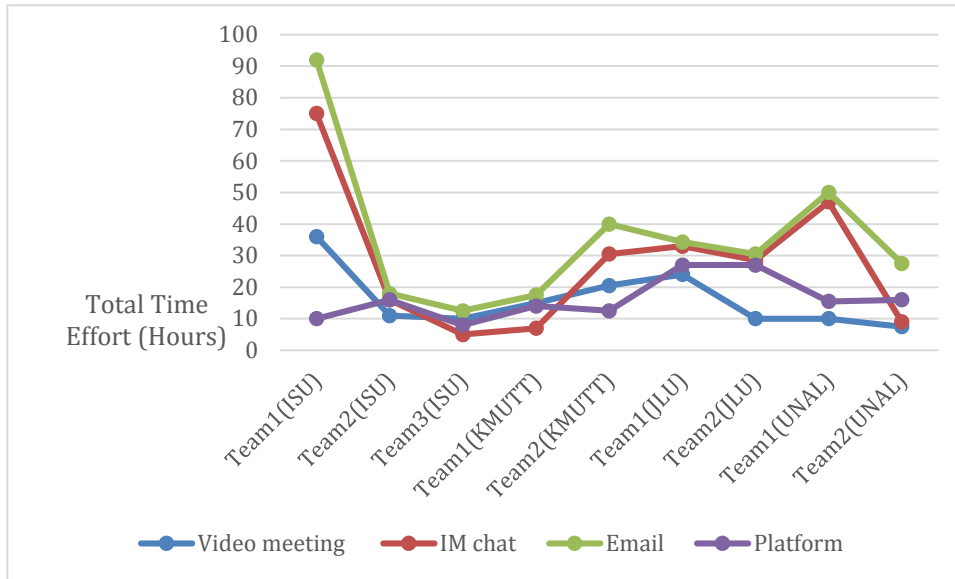
Figure 4. Total time effort of cross-site media for each team at Fall 2014

Figure 4 shows the statistics of time effort for cross-site media. It shows that teams take much more time on Email, which means Email is always the first choice for each team. However, for the second choice, some teams prefer IM chat and some teams prefer platform media. In general, video meeting is not a preference for each team. Figure 5 shows more details.

This pie chart below shows that Email occupies the largest proportion of time effort for cross-site media. We can see teams tend to choose text based communication media such as Email and IM chat.
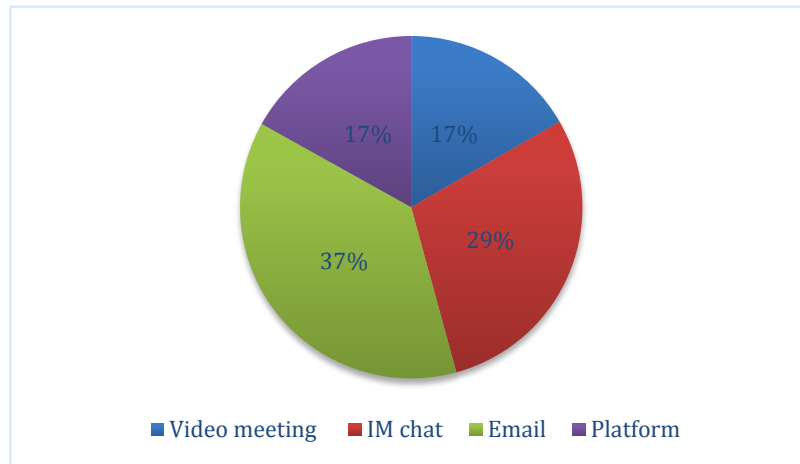
Figure 5. Proportion of time effort in cross-site media at Fall 2014
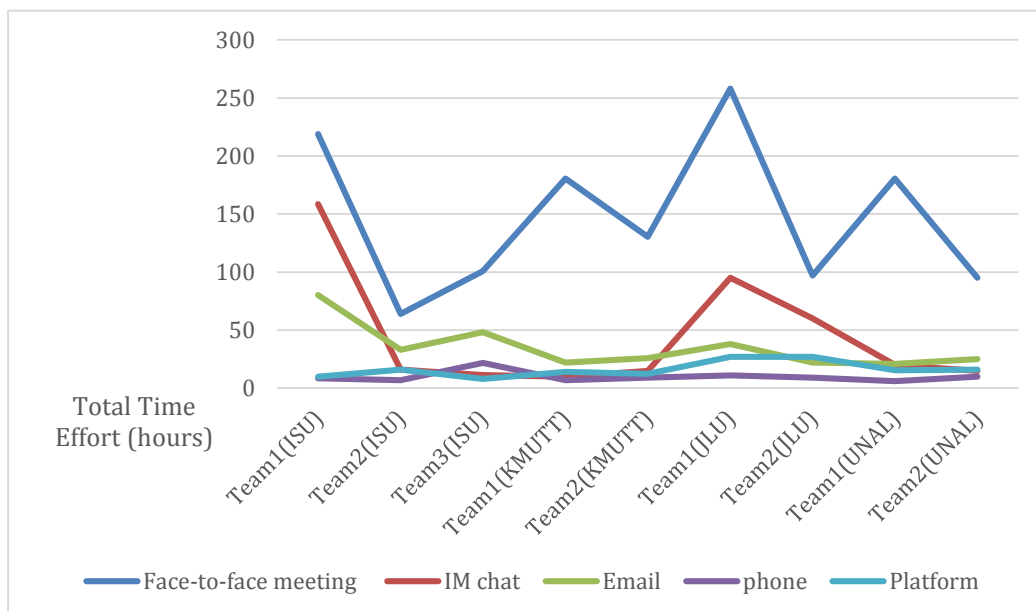
Figure 6. Total time effort of local-site media for each team at Fall 2014

As opposed to distributed communication, Figure 6 shows the communication media used in co-located development. Face-to-Face meeting occupies the biggest proportion of all media. The second choice is IM chat or Email.
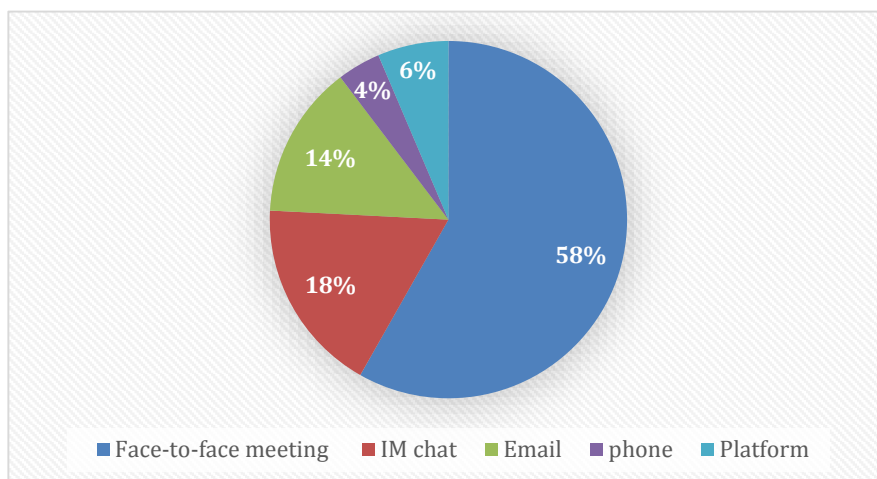
Figure 7. Proportion of time effort in local site media at Fall 2014

Figure 7 shows that teams prefer face-to-face meetings in co-located development. However, it is still hard to see any preference in the choice of audio based media. We can see that phone occupies the smallest proportion in all media. Phone is actually a more direct and easy communication media at collocated sites. Why do people not like to choose phone as communication media with each other at distributed sites? A developer at one team in ISU commented on this issue in an interview, as follows.

"We usually don't use phone although it is direct to use and doesn't have big connection issues like Skype. Phone is always used to contact to make sure of a meeting time or location. It's hard to solve a specific problem through phone. We still prefer to solve some problems face to face if we are located at the same place."

We can see phone is still not a good way to be used for co-located development, but it is helpful to set up face-to-face meetings.

For this hypothesis, we can see people tend to use Email and prefer text based communication media. Because the teams in different sites may not have the same native language or same culture, it is hard to communicate with other sites through video and audio. People usually spend too much time on understanding the original purpose of the people at another site. Sometimes, speaking communication even may cause misunderstandings and delay projects. In addition, the internet connection is also a big obstacle to use video and audio based tools. The quality of user experience depends on the Internet connection quality. Therefore, text based tools such as Email or IM chat are efficient communication media to use on DSD. On co-located development, we can see people tend to use face-to-face as their main communion media. However, it is not clear to see any preference tends on video or audio based media.

Hypothesis 2: Culture difference does have an effect on the choice of communication media.

We know culture difference is a big obstacle for communication in DSD **[4]**. In general. American culture is very task oriented, and the size of the development team for one module is usually small. However, team size in China and India is usually much bigger because of cheap human resources.   Asian as well as Southern and Eastern Europeans value personal relationship more than the task on hand **[10]**. In Asia and Europe, people respect authority more than American. Managers in Asia and

Europe have strong execution without too much concerns from lower-level employee. American managers have to discuss with their lower level employees most decisions and comments **[11]**.

We wonder if culture difference has effects on the choice of communication media. We know culture differences can include team culture difference, country culture difference and site culture difference. Here we just consider country culture difference.

From the hypothesis 1 result, we know email is the first media choice in distributed sites and face-to-face meeting is the first media choice in co-located development. However, for choosing other communication media, the percentage depends on teams. If we observe the team member composition, we found the team composed of Chinese members usually choose IM chat as their second choice. We can't, of course, rule out non-cultural causality, although two semesters' data seems to be good. However, after talking with these teams, we found that it is a common use habit to use IM in China. We believe that this result is affected by Chinese culture in some way.

We can see from Figure 4 and Figure 6, team1 and team2 in JLU and team1 in ISU choose IM chat more frequently than other teams. If we observe the team composition of these teams, they are all composed of Chinese team members. In case this is a coincidence, we also can see the data in Figure 8 for Fall 2013.
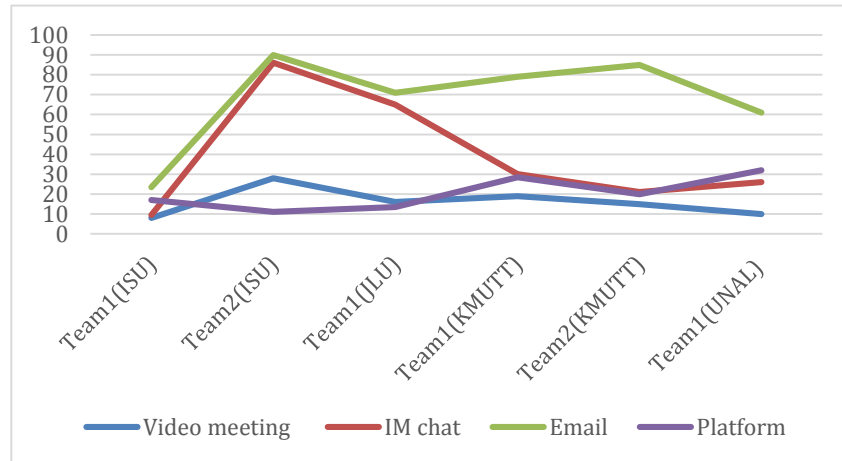
Figure 8 Total time effort of cross-site media for each team at Fall 2013
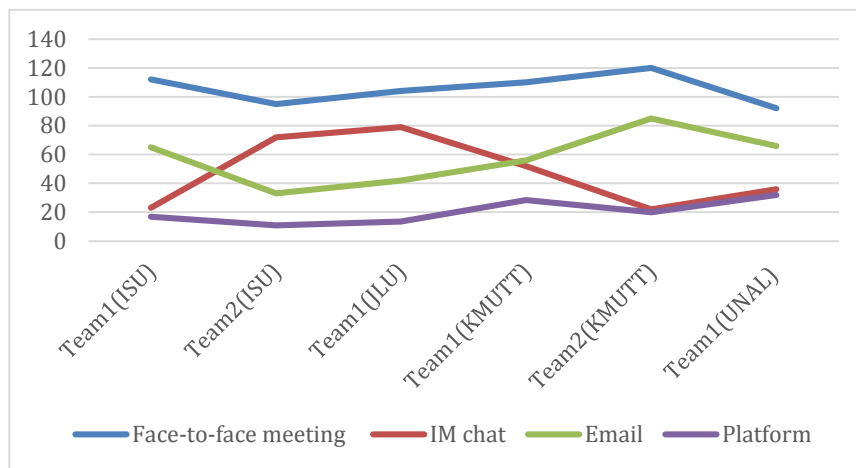
Figure 9. Total time effort of local-site media for each team at Fall 2013

Figure 8 and Figure 9 show the time effort for each team at 2013. As we propose, Team 2 in ISU and team1 in JLU prefer IM chat compared with other teams. Here Team 2 in ISU and team1 in JLU are composed of Chinese team members. This situation is the same as the data at Fall 2014. In China, people prefer to use IM in social life. [14]

Based on different countries' policies, people may not be allowed to use some communication tools. For example, in China, all Google services are blocked. In the

DSD project we studied, when all teams discussed choosing communication tools, all communication tools used by Google services were not considered. The teams in JLU can't use Google services because of the policies. The country culture difference here has an important effect on choosing communication media.

Hypothesis 3: Media synchronicity theory is helpful to select communication tool for DSD projects. [Media synchronicity theory, A.R.Dennis [12]]

When we think how to choose communication tools in DSD projects, does media synchronicity theory help? A.R.Dennis proposed a media synchronicity theory **[12]**, which gives five characteristics for a communication tool. These five characteristics are immediacy of feedback, symbol variety, parallelism, rehearsability and reprocessability.

(1) Immediacy of feedback: This characteristic shows that the communication tool could let the sender know if the message has been received, understood and acted upon by recipients. For example, you will see the received note when recipients receive the message on imessage (Apple's message service).

(2) Symbol variety: This characteristic shows that communication media can express various messages and meanings. For example, people can use IM to send emoticons to express emotion.

(3) Parallelism: This characteristic shows that communication tools can deliver the message synchronously, which means the recipients can receive the message at

the same time as the sender sends message.

(4) Rehearsability: This characteristic allows the sender to review, rethink and rephrase a message before it is sent to recipients. For example, senders of emails are allowed to review contents by senders before the message is sent.

(5) Reprocessability. This characteristic allows the sender and recipients to review, rethink and rephrase after the messages are sent. Senders are still allowed to review contents after the message is sent.

Tuomas at al proposed that these five characteristics can be classified into two directions: sharing information and building common understanding **[9]**. Tuomas at al also think sharing information can include Parallelism, Rehearsability and Reprocessability. Building common understanding can include Immediacy and Symbol variety characteristics. The communication tools with a higher level of sharing information can handle simple and concise tasks. The communication tools with a higher level of building common understanding can handle uncertain and ambiguous tasks. Based on our understanding and discussion, we analyzed the main five communication media we used in the DSD project we studied.
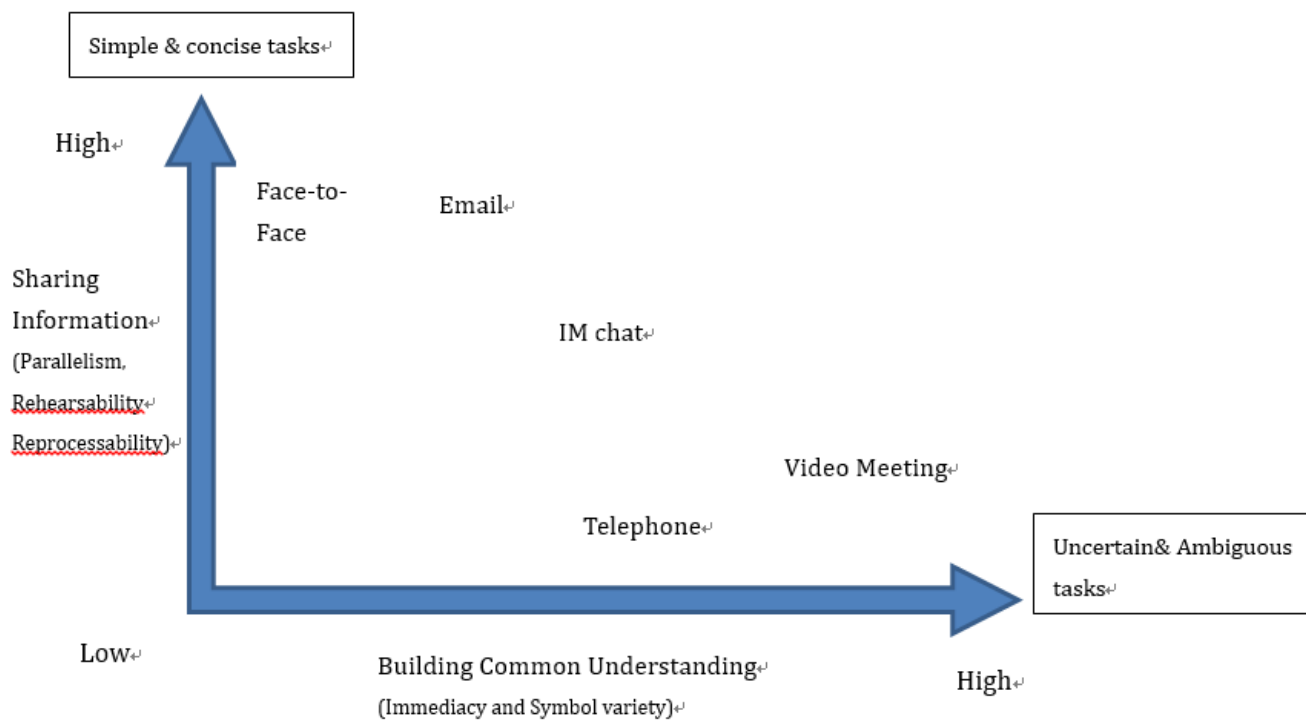
Figure 10 The media synchronicity of communication media

In Figure 10, we summarize these five main communication media in Media Synchronicity. In our analysis, Face-to-Face meeting, Email, IM chat and video meeting correspond with figure 10 and are consistent with media synchronicity theory. However, telephone has very low efficient performance during the studied project. In media synchronicity theory, telephone should have a higher sharing information level than video meeting and more concise and simple tasks. Actually, people rarely use telephones as their communication media at cross-country sites because of the expensive cost of international phone calls services. In co-located sites, people only use telephones to confirm the location and time of a meeting. Also, it's hard to describe a concrete problem and solve a concise task via telephone. As internet technology developed, people tended to use some IM tools and video tools instead of the traditional telephone media. Therefore, telephone has a worse performance in

either sharing information or building common understanding. We believe telephone probably will be displaced as a traditional communication media as technologies continue to develop in the future.

Hypothesis 4: The number of team members does not have effect on the total time effort of project.

People resources is a key factor that affects the process of a project. What we are interested in is what role the size of a team plays in distributed software development. Does the number of team members affect the time effort of a project? What other possible factors could affect the time effort of a project?

Brooks claimed that "Adding manpower to a late project makes it later" in his 1975 book The Mythical Man-Month **[13].** Even when hiring an experienced technical expert in a related field, the new hire still needs time to become familiar with the people whom he or she will cooperate with and the project. Sometimes you may also need to redesign the module structure or task structure so that everyone has something to do. Therefore, Brooks added "Nine women can't make a baby in one month" in his book.

Figure 11. People hours and human distribution on Fall 2014

In our case study, the manpower we considered is the number of team members. We have 9 teams in the case study Fall 2014. The detail of human distribution is shown in Table 1. We compare the human distribution and the people hours for each team on this project.

Figure 11 shows there is no obvious relationship between these two factors. The team with the most human resource is team 1 in JLU, but it does not have the biggest effort. Similarly, the teams with least human resources, team 1 and team 2 in UNAL, both have average time effort (people hours) on the project. Therefore, we consider human distribution is not an important factor to impact the time efforts on the project if the team size is much more than the necessary team size. However, if the team size is less than the necessary team size, the project would not proceed successfully. Determining the necessary team size is a topic for separate investigation, and has been much explored for co-located teams. Besides the human distribution factor, we also may consider the module structure distribution as a possible factor to impact the effort of teams as a future topic.

3.2 Communication issues

One important goal we tried to investigate is communication issues in DSD. We tracked all communication issues recorded in our data set and found all issues we found have been included in these categories in table 3.

Table 3. Communication issues types

Type	Description	Solution media	Issue No. (example, not all)
Artifacts design & definition	Artifacts are defined as the output during the development process such as requirements, data structure, codes and test plans. Teams are not clear about the requirements or plans of their artifacts. Teams need communication to make sure their scope of work.	Face-to face, IM	K_C1 K_S2 J_K2 J_D3
Artifacts change	One team changes the contents of their artifacts. Other teams related with this module spend time receiving notification and solve the consequences that the changes caused.	IM, email	K_C12 I_A2 J_K16
Texts miscommunication	During review process within the teams, sometimes it became difficult to understand writer's purpose. People even misunderstood the documents while presenting some concepts within the document.	Face-to-face. Video meeting, Phone	I_O1 I_A3 K_C14 K_S1 I_A4
Language miscommunication	People can't communicate or misunderstand others because of speaking accents or cultural difference.	IM text	I_A_02
Receiving Notification delay	Message is missed or not received in time because of tools problems or receivers did not check massages in time.	Video meeting, IM,	I_O3 I_O4 I_A1 I_A7
People leave	The team member leaves the team and other backup people have to take charge of the work. All communication should be recorded for reviewing.	Email, Face-to-Face	I_E2 I_E3

We found that there are six main types of communication issues including requirements definition, requirement change, texts miscommunication, language miscommunication, receiving notification delay and people leave. Table 3 describes these problems.

(1) **Artifacts design and definition.** This type of issue usually happens at the beginning of the project. Teams are not very clear about the requirements for their artifacts, so they need to communicate with other teams and instructors in details. This kind of communication should be detail-task oriented and straight forward. Therefore, IM and face-to-face meetings are good solutions for this kind of issue. One example of this kind of issue is issue K_C1. The description and solution of issues recorded in the form is listed below:

Description: "It seems like KMUTT-team1 and KMUTT-team2 don't understand the same scope of work. We need to clarify the requirement of our modules"

Solution: "Teams have a lot of talk in Line and Facebook. Conduct urgent meeting to agree scope of work again, also include professor O."

(2) **Artifacts change.** This type of issue is a critical issue for projects, which usually cost teams much effort in fixing the requirements change. For this kind of issue, team members need to understand changing contents well and are able to track documents easily. Therefore, the solution is done through text based communication tools, like IM and Email. One example of this kind of issue is ISU_O2.

Description: "We have made requests to their database team previously. Currently we revised our design and found that the changes requested previously are not required."

Solution: "We have mailed them saying that the changes are not required."

(3) **Text miscommunication**. If team members only read documents to understand tasks without other communication, it is easy to misunderstand the original purpose of writers, which would cost much time to find the miscommunication and fix problems. This kind of issue needs to be solved by audio based or video based communication, so task holders would have a discussion to understand the document in details. One example is listed below:

Description: "Server Team in China asked if there are any authentication between Facilitator and Server. When he first sent the Email, I forwarded it to all of our members, but I didn't get any response. One reason was we were busy preparing for the mid-term exams, and the other reason was all of us didn't understand what his meaning is. It caused some delay in the communication between Server and Facilitator."

Solution: "After another Email, we suggested them to use IM Chat - QQ to have a discussion. We had a group chat with Server Team in QQ and it's more convenient for us to discuss in time. Sometimes it is forgotten to receive emails easily. After discussion in QQ, we solved this problem effectively."

(4) **Language miscommunication**. This issue usually happens between two teams with different native languages and culture. Although teams all speak English during video and face-to-face meetings, this kind of problem still exists because of accents and culture differences. These issues usually are solved by IM. They can get responses in time on IM and it is text based which is easy to understand. One example is ISU_A2.

Description: "Our team members tried to have a video meeting with the PM of Team1 in K University, talking about the functions of Face Mapper Module, the communication methods between our Facilitator and Face Mapper, and data format to transmission. But we can't communicate with each other because of accents and poor English. English is not our native language. We both had some difficulties in expressing our own ideas and understanding theirs."

Solution: "We changed video meeting into text based meeting at last, and use Skype text to each other about the ideas and understanding. It's much easier for us to communicate through texting."

(5) **Receiving notification delay.** This issue is also a critical issue. If this issue happens, other related tasks have to be suspended and wait for responses. IM and video meeting can solve this kind of issue because notification could be delivered on time and recipients would pay more attention to the issue. One example is issue I_V2

Description: "On Oct 21, we received an email from KMUTT team2, but we

could not acknowledge in time because it went into our spam folder"

Solution: "We discussed a rule that if recipients don't respond within 24 hours, sender would use IM to notify recipients."

(6) **People leave**. This issue also usually happens in an industrial environment. Some students dropped this course or left the team temporarily because of personal reasons. This type of issue is solved by multiple communication tools. Within teams, backup team members had to communicate with team members and be familiar with previous tasks as soon as possible. The communication between teams is done through text based communication tools such as email. Backup people could review and track previous tasks through previous email records. One example is ISU_E1

Description: "UNAL team2 needed to update their API and this was communicated a while ago, due to change in developers and time to get acquainted to the process this took some extra time and remained in the pipeline"

Solution: "Good knowledge transition from previous to new developer"

Based on the communication issues and solutions we saw, we think such solutions would be helpful to improve the efficiency of DSD, especially reducing the time cost of communication. However, this strategy should be verified in industrial projects in the future.

CHAPTER IV

CHALLENGES

In this chapter, we will analyze the challenges we encountered during the study. What we focus on are the difficulties to get the right data, and the challenges to verify the hypothesis. Although there are some challenges during this study, we tried to overcome the challenges and controlled them under minor risks.

4.1 The challenges to get the right data

During the process of collecting data, we faced different kinds of problems that affect the accuracy of data.

4.1.1   Timeliness

One important challenge we faced is how to ensure the data can be collected in time. As mentioned before, we collected data weekly through data collection forms. Every team member records his/her data once a week, since one week is short enough for them to recall the exact time effort for that week. However, in fact we still cannot ensure the timeliness in the process. One big problem is that PMs usually delayed or forgot to send the teams' data collection forms for the Fall 2013 data set. Every team member still had a strong motivation to submit his/her data collection form to their PM because of course requirements. For the convenience of submission, we started to use Google doc to collect data at Fall 2014. The data of every team member can be viewed and tracked on Google doc in real time, which greatly reduces delays. If any

data seem to be abnormal or strange, it is also convenient to contact directly the person who filled out the form.

4.1.2 Completeness

Completeness is also an important challenge in this study. Missing data would have significant effects on our verification or might cause incorrect conclusions. To ensure the completeness of data, we checked the data every week and tried to have all team members complete the form. Some reasons for missing data were the holidays or exams during which time students did not expend any effort on the project. What we still cannot solve is that they could not recall the missing data when they delayed submitting the form by several weeks. Sometimes they may just made up some vague or nonstandard data. We only see these data as invalid data. Fortunately, these missing data form only a small percentage (about 11%) in total, and probably have only a small effect on our analysis.

4.1.3 Duplication

Duplication of data, especially the duplication of communication issues may raise the difficulty of verification and investigation of issues. When we collected the data at Fall 2013, we found that some same communication issues seemed to be recorded in different team members' form. We cannot tell if these persons are involved in the same issues. It was also hard for them to recall these issues when we found the problem. Therefore, to reduce the duplication of communication issues, we recorded

the communication in one sheet within one team, and each issue has unique issue id. Before providing the data, students must check if the issue has been recorded. If it was recorded, students need to check if they have any supplemental information about the issues. Therefore, Fall 2014 data are much clearer and does not have any duplication.

CHAPTER V

CONCLUSION AND FUTURE WORKS

In our study, we recorded the time effort of a project, the use of communication tools and communication issues during two semesters' of the DSD course across four countries. We presented an analysis of these data showing a relationship between country culture and the choice of communication media, as well as a relationship between team size and time effort on the project. We believed that the diversity of country background of team members would build a good choice of communication tools. However, adding more or less manpower would not have much effects on the time effort of project development if the team size is much more than necessary size. Our analysis also helped to verify the media synchronicity theory's effect on selecting communication tools. Therefore, when one deals with simple and concise tasks, choose more direct communication media, such as face-to-face meetings. If one deals with some uncertain and ambiguous tasks, choose communication tools that make it easy to build common understanding, such as IM tools, or video meetings. Through the investigation of our data, we also found that persons tend to select text-based communication tools in DSD and select speaking based communication tools in collocated development.

We have also investigated communication issues in the development process. There are main six types of communication issues including artifacts definition, artifacts change, text misunderstanding, speaking misunderstanding, receiving

notification delay and people leave. We believed that a deep understanding of communication issues is critical for the success of today's distributed software development. Based on the solution of communication issues we encountered in the study, we also provided solutions for these six types of issues. The strategy of solving communication issues would be helpful to reduce the delay and increase the efficiency in development.

As more new technologies develop, new communication media will become available. It is necessary and important to follow and evaluate new communication tools in distributed projects. If it is possible, new advancements in this area should be introduced in the study, such as microblogging services and code collaboration with communication. We also wish there was more evidence available from industry to support our research results in this study. The data in this study is course based, just simulating the industrial environment, but it is not a real industrial software development environment. The investigation of data from industry will be more convincing about communication issues and their solutions in DSD.

CHAPTER VI

ACKNOWLEDGEMENTS

I would like to thank my committee chair and advisor, Professor David Weiss, and my committee members, Professor Carl K. Chang and Professor Les Miller for their guidance and support throughout the course of this research. I really appreciate the help from my respectful Professor, Dr. David Weiss. He guided my research and revised my thesis for several versions. I am very touched and motivated by his hard working and spirits. Without his help, I will not finish this study and thesis.

In addition, I would also like to thank my friends, colleagues, the department faculty and staff for making my time at Iowa State University a wonderful experience. I want to also offer my appreciation to those students who have taken coms510 and were willing to participate in my study, without whom, this thesis would not have been possible.

Finally, thanks to my family for their encouragement, patience, respect and love.

REFERENCES

[1] J.D Herbsleb and D. Moitra (2001), "Global Software Development, IEEE Software, March/April, USA, p. 16-20,

[2] R. D. Battin, R. Crocker, J. Kreidler, and K. Subramanian. Leveraging resources in global software development. IEEE Software, 18(2):70–77, March/April 2001.

[3] E. Carmel and R. Agarwal. Tactical approaches for alleviating distance in global software development. IEEE Software, 2(18):22–29, March/April 2001.

[4] G. M. Olson and J. S. Olson. Distance matters. Human- Computer Interaction, 15(2/3):139–178, 2000.

[5] J. Herbsleb and A. Mockus. An empirical study of speed and communication in globally distributed software development. IEEE Transactions on Software Engineering, 29(6):481–94, 2003.

[6] Christian Bird, Nachiappan N. Does distributed development affect software quality? An empirical case study of windows vista, 31st international conference on software engineering, IEEE, 2009

[7] D.E. Perry, N.A. Staudenmayer, and L.G. Votta, "People, Organizations, and Process Improvement," IEEE Software, vol. 11, no. 4, pp. 36-45, 1994.

[8] R. L. Daft and R. H. Lengel, "Information richness: a new approach to managerial behavior and organizational design," in Research in Organizational Behavior, B. Staw and L. Cummings, Eds. JAI Press, Greenwich, Connecticut, 1984, pp. 191–233.

[9] Tuomas N. Arttu Piri, "Reflecting the choice and usage of communication tools in DSD projects with media synchronicity theory", 2010 International conference on Global Software Engineering, 2010

[10] Hall, E. T.,&Hall, M. R. *Understanding cultural differences: Germans, French and Americans.* Yarmouth, ME: Intercultural Press, 1990

[11] Hofstede, G. *Cultures and organizations: Software of the mind.* London: McGraw-Hill, 1991.

[12] A. R. Dennis and J. S. Valacich, "*Rethinking media richness: Towards a theory of media synchronicity,*" in Proceedings of the 32nd Hawaii International Conference on System Sciences, 1999.

[13] Frederick P. Brooks, Jr. *"The Mythical Man-Month"*. 1995 [1975]. Addison-Wesley

[14] Li Zheng, Bin Chen, 2014 Chinese instant message use habit analysis and report, 2014, Beijing

[15] R.E. Kraut and L.A. Streeter, "Coordination in Software Development,"Comm. the ACM, vol. 38, no. 3, pp. 69-81, 1995.

[16] Basili, Victor R., and David M. Weiss. "A methodology for collecting valid software engineering data." Software Engineering, IEEE Transactions on 6 (1984): 728-738.

[17] T.J. Allen, Managing the Flow of Technology. MIT Press, 1977.

APPENDIX A

DATA COLLECTION FORM

Data collection sheet form
(part 1)

Date	Week No.	Name	Role	Total Effort Expended (hours)	Interface Effort (hours)			
					Issue1 (options)	Effort(hours)	Issue 2 (options)	Effort(hours)
9/25/13	1	LI	PM(Liaison)	14	All modules	1	Client/Databas	1
9/25/13	1	Kevin	Architect	7	Client/Database	1.5		
9/27/13	1	Jack	Developer	15.5	All modules	1		

Data collection sheet
form (part 2)

	Comunication Tools Effort (hours)										Communication Issues	
Total number of communication tools involved	Face to Face local chat (hours)	video meeting (hours)	IM chat sessions Number/effort (hours)		phone chat numbers/effort (hours)	Email Number/effort (hours)		platform	Other (csite/local)	Issue ID (SchoolName_Team Name_ID)	Effort to resolve (hours)
			cross site	local site		Cross-site	Local-site				
12	6.5	0	0.5	1	2/1	6/0.5	7/0.5	1.5		jlu_DBTiger_01	1
5.5	3	0				5/0.5		0.5		jlu_DBTiger_01	0.5
	8					1/0.2		1		jlu_DBTiger_02	1

Issue Form

Miscommunication ID	Face to Face local chat	what communication tools involved in miscommunication issue									Issue Description, including description of misunderstanding or incorrect assumption(s) made, error(s) that resulted.	Solution description, including how to solve problem(s) and the way you take.	Comments
---------------------	-------------------------	video meeting		IM		Phone	Email		Platform				
		options	others	options	others		options	others	options	others			
jlu_DBTiger_01	Yes	Google hangout		Skype		Yes	Gmail		Assembla		ISU team has changed the requirement of server interface-test deadline. There are only 3 days before first test deadline. ...	We set up a video meeting to negotiate this issue with ISU team. And they agree with the changing of interface 1.3 and 2.6.	They should inform us of this requirment changing through ...

APPENDIX B

GQM

GQM for Distributed Software Development Course

Following is the goals and associated questions and metrics for the DSD course for Fall 2013

	Goal I – Identify methods for teaching effective distributed software development	
Questions	**Metrics**	**Where collected**
a. Did the project permit and encourage a distributed organization?	4. Effort to agree on cross-site interfaces, both total and by module.	Weekly data spreadsheet, columns C,D,E,F
b. Were the students able to modularize effectively both within and across teams?	5. Effort expended by each team and each team member.	Communication effort by week on weekly data spreadsheet (doesn't measure total effort, however)
c. How much difficulty did the students have in establishing cross-team interfaces?	8. Number of defects found in final result, especially critical defects.	Test results (form for reporting not yet determined, but should be part of V&V plans), also communication issues, particularly those that resulted in critical defect in final result (if any)
d. Was the work evenly divided both within a team and between teams, or did just one or two students on each team do all the work?		
e. What intersite communication methods were used?		
f. How effective were the intersite communication methods?		
g. What were the key ideas distinguishing distributed development from co-located development that the students learned? How successful were the students in completing the project?		

Goal II – Identify key techniques that enable effective distributed software development.		
Questions	**Metrics**	**Where collected**
h. Did the definition of architecture as a set of binary relations clarify and identify independent work assignments (modules) and the dependencies among them?	16. Channel (face-to-face, video conference, email, chat, other), Parties(team, person), Type(local or global), Reason for communication, Current Activity(Introduction, Planning, Requirement, Design, Development, Testing, other)	Weekly data spreadsheet, columns G through J, but reason for communication not included on spreadsheet (collected elsewhere?)
i. Did identifying liaisons make a difference?	17. The number of hours spent in communications by each person in each activity	Weekly data spreadsheet, columns G through , but activity not included on spreadsheet.
j. How effective were the communication tools?		
k. Which teams were most effective and why?		
l. How did the quality of the resulting product compare to the same product when developed by a co-located team?		
m. How did the time to develop compare to the same project when developed by a co-located team?		
n. Does the actual communications followed the expected communication model in DSD?		
o. Does the actual communications follow the expected communication model in DSD?		
p. What is the effort spent in communication comparing to other activities (e.g. planning, developing artifacts in each phase)?		
q. Has communication difficulty been effectively handled?		