International Journal of Software Engineering and Its Applications
Vol.8, No.6 (2014), pp.241-252
http://dx.doi.org/10.14257/ijseia.2014.8.6.19

How Geographic Distribution Affects Development Organizations: A Survey on Communication between Developers*

Jongdae Han[1] and Woosung Jung[2†]

Department of Digital Software, Sangmyung University
20, Hongimun 2-GIL, Jongno-gu, Seoul, Korea
Department of Computer Engineering, Chungbuk National University
52 Naesudong-ro, Heungdeok-gu, Cheongju, Korea
[1]elvenwhite@smu.ac.kr, [2]wsjung@cbnu.ac.kr

Abstract

Geographically distributed software development (GDSD) is popular due to various advantages. However, GDSD has temporal, physical and socio-cultural challenges related to distance between developers. In this paper, we measure how geographic/physical distance affects development organizations through a survey.

Keywords: *Geographically distributed software development, communication structure, social network, team management*

1. Introduction

The size of most software has increased over the years. Because software development is a team activity, a "chosen few" developers cannot take care of every complex aspect of development. The team must be well structured and organized in order to support the organization's expansion.

The organization of software development can be explained through various perspectives. As with other common business organizations, they can be considered hierarchical based on the roles and ranks of their members. They can also be regarded as object-driven organizations that exist to achieve a single goal, like building up and maintaining specific software. Furthermore, team organizations can assign functionalities to the team members and encourage collaboration between them.

Geographically distributed software development (GDSD) organization can be distinguished from other software development organizations due to the restricted communication channels. Many researchers pointed out [2-5] that GDSD suffers from decreased productivity due to cultural / temporal issues that cause problems with communication among developers. Therefore, development organizations should be reviewed in terms of communication.

Our research goal is to determine how geographic distribution affects development organization. Geographic distribution is assumed to hinder forging social network between developers, weaken teamwork within the organization, threaten collaborative structure, and force developer to use expensive communication channel.

In this paper, we surveyed developers who work for GDSD firms in order to analyze how the geographic distribution affects the developer social network. The rest of paper

* This paper has been extended from our former study [1]
† Corresponding author

ISSN:1738-9984 IJSEIA
Copyright ⓒ 2014 SERSC

is organized as follows: In Section 2, current studies on GDSD and other relevant background are introduced. In Section 3, survey results and its analysis are provided and we wrap up our findings and discussion in Section 4.

2. Related Works

Herbsleb [2] analyzed project properties including global dispersion, number of developers, severity of problem, *etc*. that affect the delay between modification request and resolution. He performed a multiple regression to find that the geographic distribution tends to increase the development organization's communication cost and size.

Furthermore, Ngamkajornwiwat [6] studied the development ecosphere of open source software. In his study, he categorized the participants of open source software projects and observed how the open source community evolves by participating in multiple projects. He states that participants tend to lose interest in projects proportional to the age of the project. The communication patterns between categorized developers can be discovered through this research.

Cataldo[7,8] researches about congruency within GDSD organizations. He measures the congruency factor of four aspects: structural congruence, geographic congruence, modification request congruence, and internet relay chatting (IRC) congruence. The research shows that geographic distribution acts as an obstacle against team congruency, but proper tools can reduce the damage.

In addition, Zhou's research [9] quantifies the communication costs of GDSD organizations. The Project Communication Model, which is proposed by the research, insists that the project's size and complexity, communication channel efficiency, language barrier, domain/technical knowledge between distributed teams, and number of participants derive the sum of the costs.

Conchúir's research [10] shows many significant beneficial aspects of GDSD as well as associated risks and challenges. His qualitative interviews show that there are various issues to the assumed benefits like reduced costs, leveraging time-zone effectiveness, cross-site modularization, access to larger labor pool, innovation, shared best practice, and closer proximity to market. Those issues are considered to build up our questionnaire.

Nagappan [11] applies eight metrics to review how organizational structure affect software quality. The metrics are Number of Engineers, Number of Ex-Engineers, Edit Frequency, Depth of Master Ownership, Percentage of Org contributing to development, Level of Organizational Code Ownership, Overall Organization Ownership, and Organization Intersection Factor. The research empirically tests relation between those metrics and code quality metrics, and shows that there is clear correlation between organizational structure and software failure-proneness. Although the research is not strictly concerning geographically distributed software system, it assumes the development organizations are consisted with many teams of individuals working together. Therefore the research give several useful insights about organizational structure for our research.

3. Survey Results and Analysis

Our questionnaire is an extended version of Herbsleb's questionnaire [2]. The answers use the Likert scale, with several exceptions for open-ended questions. 24 participants who work for Korean GDSD organizations answered the questionnaire.

 Copyright ⓒ 2014 SERSC

The survey participants work for medium- to large-sized commercial software firms. The development fields include web applications, PC online games, and smartphone applications. The development sites that participated in our survey are mostly domestic with little cultural/linguistic gaps, because most of their colleagues are also Korean.

3.1. Survey Goal

Our preliminary research yields the following goals for the survey. On the GDSD environment:

- Do developers tend to be 'socially' closer to their collocated colleagues than separated counterparts?
- Does geographic distribution hinder teamwork of the organizations?
- How does geographic distribution hurt the collaborative structure of organizations?
- How does geographic distribution disturb sharing information and expertise within the organizations?
- How does geographic distribution force developers to re-configure their communication channels? Is the cost of channels affected by distribution?

For these goals, we set up thirty-eight questions that can be found at https://space.zeo.net/g/i58b. We also conducted post-survey non-constructive interviews with several participants and other developers to get additional knowledge of the survey results.

3.2. Results and Analysis

3.2.1. Social Network on GDSD Environment: The social network established between developers acts as an informal communication structure and helps collaboration [8, 12]. The social network is expressed in terms of reliability, trustworthiness, personal intimacy, and comfort. Six questions* are asked to see if geographic distribution affects those emotional traits between developers. The participant answers are tested by the Cronbach Alpha method to check the internal consistency reliability. Table 1 shows the result. The results from the test result in α values of 0.753 and 0.783, respectively. Therefore, we can safely assume our questions are internally consistent.

Table 2 shows the analysis results that determine how geographic distribution affects the network. The mean Likert scale score for the collocated group is 2.07, whereas that of distributed group is 2.93. This means developers agree that they tend to form stronger social networks with colleagues in the same physical site. Q7 and Q12 yield the most distinct results, with mean Likert scale score discrepancies at 1.33 and 1.25, respectively. Q7 and Q12 question the strength of the social network that is not expressively related with their work. Therefore, geographic distribution can be said to hurt the informal social network established by developers more than a formal, work-related one. Q8 that asks in-work reliability had the least distinct discrepancy, and also supports our analysis.

* Q7, 8, 9, 10, 11 and 12

Copyright ⓒ 2014SERSC

Table 1. Internal Consistency Test for Social Networks on GDSD Environment

Reliability Statistics:Collocated **Reliability Statistics:Distributed**

Cronbach α	N of items	Cronbach α	N of items
.753	6	.783	6

Item-Total Statistics

	Cronbach α if item deleted		Cronbach α if item deleted
Q71	.776	Q72	.767
Q81	.722	Q82	.769
Q91	.683	Q92	.743
Q101	.715	Q102	.714
Q111	.657	Q112	.746
Q121	.734	Q122	.755

Table 2. Difference of Social Network Strength by Geographic Distribution

Strength of Social Network:Collocated

		Responses		Percent of Cases
		N	Percent	
Collocated[a]	Strongly agree	28	19.4%	116.7%
	Agree	86	59.7%	358.3%
	Neutral	22	15.3%	91.7%
	Disagree	8	5.6%	33.3%
Total		144	100.0%	600.0%

a. Response group

Strength of Social Network:Distributed

		Responses		Percent of Cases
		N	Percent	
Distributed[a]	Strongly agree	9	6.3%	37.5%
	Agree	46	31.9%	191.7%
	Neutral	46	31.9%	191.7%
	Disagree	32	22.2%	133.3%
	Strongly disagree	11	7.6%	45.8%
Total		144	100.0%	600.0%

a. Response group

Copyright ⓒ 2014 SERSC

3.2.2. Teamwork on GDSD Environment: Teamwork is defined as degree of collaboration needed to achieve team-wide goals. It manifests itself by sharing responsibility, voluntary cooperation, receptiveness between members, etc. Six questions* find how the GDSD environment affects teamwork within collocated / distributed teams. The participants' answers are also tested by the Cronbach Alpha method to test internal consistency reliability. Table 3 shows the results, where Q37 decreases the α value of the answers for collocated colleagues from 0.749 to 0.709. Although Q37 asks about how a developer feels about his teammates' skillset, we decided to drop it to maintain consistency. Furthermore, the α values of the result are 0.749 and 0.873, respectively, which are high enough to ensure internal consistency.

Table 3. Internal Consistency Test for Teamwork on GDSD Environment

Reliability Statistics:Collocated **Reliability Statistics:Distributed**

Cronbach α	N of items	Cronbach α	N of items
.709	6	.873	5

Item-Total Statistics

	Cronbach α if item deleted		Cronbach α if item deleted
Q131	.654	Q132	.886
Q141	.690	Q142	.815
Q151	.644	Q152	.842
Q161	.609	Q162	.822
Q171	.653	Q172	.854
Q37	.749		

Table 4 shows the analysis results for five questions about teamwork issues in the GDSD environment. The teamwork for a geographically distributed team is notably weaker than that of a collocated team. The mean Likert scale score for the collocated group is 2.10, whereas that of the distributed group is 3.09. 73.3% of the developers agree that teamwork exists with collocated colleagues, while only 32.5% agree for colleagues at distant locations. This is consistent for both inbound and outbound teamwork. Therefore, long-range geographic distribution hurts teamwork between developers. If an organization adopts methodologies that require hardened teamwork, the members of a team should be located in the same location.

3.2.3. Collaborative Structure on GDSD Environment: A collaborative structure refers to the structural allocations of helpful communication channels that support members of the organization. The structure helps in the planning and scheduling of jobs, sharing expertise, and effectively synchronizing job progress. The communications cost between the distant channels are quite "expensive," so knowing the factors required for geographically distributed active channels is essential. There are eight questions† to

* Q13, 14, 15, 16, 17 and 37

† Q18, 19, 20, 21, 22, 23, 24 and 25

Copyright ⓒ 2014SERSC

figure out these factors. These questions find the correlation between cooperative structures within a GDSD organization and communication requirement. The internal consistency test is omitted, except questions* that ask how strong developers rely on certain mediums to communicate over geographic distribution remain. Table 5 shows the results from the test†.

Table 4. Difference of Teamwork Strength by Geographical Distribution

Teamwork : Collocated

		Responses		Percent of Cases
		N	Percent	
Collocated[a]	Strongly agree	30	25.0%	125.0%
	Agree	58	48.3%	241.7%
	Neutral	22	18.3%	91.7%
	Disagree	10	8.3%	41.7%
Total		120	100.0%	500.0%

a. Response group

Teamwork : Distributed

		Responses		Percent of Cases
		N	Percent	
Distributed[a]	Strongly agree	5	4.2%	20.8%
	Agree	34	28.3%	141.7%
	Neutral	38	31.7%	158.3%
	Disagree	31	25.8%	129.2%
	Strongly disagree	12	10.0%	50.0%
Total		120	100.0%	500.0%

a. Response group

Because Q23-2 severely decreases the α value from 0.712 to 0.615, the answers are excluded for further analysis. A post-survey interview suggests that some of the participants misread the question.

Table 5. Internal Consistency Test for Collaborative Structure on GDSD

Reliability Statistics

Cronbach α	N of items
.615	4

Item-Total Statistics

	Cronbach α if item deleted
Q222	.468
reversedQ232	.712
Q24	.449
Q25	.517

* Q22-2, 23-2, 24 and 25

† Q23-2 is a negative statement, hence answers for the question is converted negatively

Copyright ⓒ 2014 SERSC

Table 6 Regression Analysis for Factors Requires Communication Channels

Coefficients

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	.492	.733		.671	.510
Q18_2	.738	.141	.732	5.249	.000
Q19_1	.320	.136	.342	2.355	.029
Q19_2	-.484	.146	-.487	-3.305	.004
Q20_1(R)	.214	.119	.253	1.796	.088

Environment

We perform a backward-elimination regression test to find the answer for the following research question:

Which organizational action causes the organization to set up communication channels over the geographic distribution?

Organization decisions manifest themselves from the answers of Q18, 19, 20 and 21, which ask issues around scheduling and job priority. Communication channels are drawn from mean Likert scale scores from Q22-2, 24 and 25, which ask issues around job synchronization and information sharing. Table 6 is the result of the regression test that find the most highly-correlated factor.

The results shown in Table 6 suggest that the top three most effective approaches that minimize the need for communication channels, which can be deducted from Q18-2, 19-1 and 19-2 answers are:

- Reactively delivering information needed to fit plan/requirement change request to distant sites
- Making flexible plans for jobs that need to be done in parallel for distant sites
- Making eminent plans for jobs that need to be done in a single site only

Therefore, we can assume that developers would able to reduce expensive communication channels over geographic distribution if they deliver relevant information whenever the development plan or requirement is changed, and applying appropriate planning policy for the site could be helpful.

3.2.4. Information and Expertise Sharing over Geographic Distribution: Recent technical innovations around communication channels for software development helps developers easily share necessary information. Although an electronic board, realtime messenger application, video conference call, and automated scheduling tools support many development organizations, the most effective communication channel is the "water cooler chat" method that still requires informal, face-to-face contact [2]. Therefore, geographic distribution is expected to disturb communication effectiveness

Copyright ⓒ 2014SERSC

by forcing developers to use less effective and more expensive channels. We ask six questions* regarding this issue.

Table 7 shows the result from the Cronbach Alpha test for internal consistency between four questions that use the Likert scale.

Table 7. Internal Consistency Test for Information and Expertise Sharing on GDSD Environment

Reliability Statistics:Collocated

Cronbach α	N of items
.550	4

Reliability Statistics:Distributed

Cronbach α	N of items
.713	4

Item-Total Statistics

	Cronbach α if item deleted		Cronbach α if item deleted
Q261	.411	Q262	.767
Q271	.693	Q272	.769
Q301	.287	Q302	.743
Q311	.486	Q311	.714

The results of Q27 show that the collocated site seriously decreases internal consistency from 0.693 to 0.550. Q27 asks if it is hard to contact a specific colleague whenever a developer needs to. Because we were unable to find why Q27 only hurts consistency for the answers from the collocated site, we did not exclude the answer from Q27. Table 8 shows the analysis results for these answers.

Table 8 shows the mean Likert scale score of the collocated group as 3.92, which means the developers do not suffer from communication when they are collocated. However, the distributed group's mean Likert scale is 2.95, so Developers tend to find it hard to communicate over geographic distribution.

The dispersion for the mean Likert scale score of Q26 is 1.80. The question asks how communication is difficult in the organizational and structural channels. On the other hand, the score dispersion of Q30 and 31, which asks how difficult it is to communicate with work artifact and expertise sharing, are 0.5 and 0.625, respectively. Therefore, we can assume those channels are more suitable in the GDSD environment.

* Q26, 27, 28, 29, 30 and 31

 Copyright ⓒ 2014 SERSC

Table 8. Difference of Communication Channel Efficiency by Geographical Distribution

Communication channels : Collocated

		Responses		Percent of Cases
		N	Percent	
Collocated[a]	Strongly agree	1	1.0%	4.2%
	Agree	4	4.2%	16.7%
	Neutral	17	17.7%	70.8%
	Disagree	54	56.3%	225.0%
	Strongly disagree	20	20.8%	83.3%
Total		96	100.0%	400.0%

a. Response group

Communication channels : Distributed

		Responses		Percent of Cases
		N	Percent	
Distributed[a]	Strongly agree	5	5.2%	20.8%
	Agree	32	33.3%	133.3%
	Neutral	28	29.2%	116.7%
	Disagree	25	26.0%	104.2%
	Strongly disagree	6	6.3%	25.0%
Total		96	100.0%	400.0%

a. Response group

We also looked at the average hours lost per month due to communication issues. While the collocated group lost 0.43 hours per month from 3.04 issues, the distributed group lost 4.87 hours per month from 3.92 problems. Considering that most of the participants work in Korea with uniform time zone, the time loss is expected to be greater if the distributed group works in different time zones.

4. Conclusion

We discussed various issues around how the GDSD environment affects social network, teamwork, collaborative structure, and communication channels in software development organization. Our survey shows that developers have trouble setting up a social network with colleagues in geographically distributed sites, especially in an informal network. The geographic distribution hinders the core element of teamwork for organizations in the GDSD environment. Furthermore, the regression analysis shows that managers must deliver information in change requests be sent to other sites as soon

Copyright ⓒ 2014SERSC

as possible. Communication channel and information sharing questions on the survey show that geographic distribution damages the structural and organizational communication channel, while communication channels based on work artifact are relatively resilient.

We will further analyze participant responses to cover up other issues regarding the GDSD environment in future papers.

Acknowledgements

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea(NRF), funded by the Ministry of Science, ICT & Future Planning(NRF-2012R1A1A1043769).

References

[1] J. Han and W. Jung, "A Study on Effects of Geographical Distribution for Software Organizations", Advanced Science and Technology Letters, vol. 46, (Software 2014), **(2014)** April, pp. 158-161.
[2] J. Herbsleb and A. Mockus, "An empirical study of speed and communication in globally distributed software development", Software Engineering, IEEE Transactions, vol. 29, no. 6, **(2003)** June, pp. 481-494.
[3] E. Carmel, "Thirteen assertions for globally dispersed software development research", System Sciences, 1997, Proceedings of the Thirtieth Hawaii International Conference, vol. 3, **(1997)** January, pp. 445-452.
[4] E. Carmel and R. Agarwal, "Tactical approaches for alleviating distance in global software development", Software, IEEE, vol. 18, no. 2, **(2001)** March-April, pp. 22-29.
[5] J. Munch, "Risk management in global software development projects: Challenges, solutions, and experience", Global Software Engineering Workshop (ICGSEW), 2011 Sixth IEEE International Conference, **(2011)** August, pp. 35.
[6] K. Ngamkajornwiwat, D. Zhang, A. Koru, L. Zhou and R. Nolker, "An exploratory study on the evolution of oss developer communities", Hawaii International Conference on System Sciences, Proceedings of the 41st Annual, **(2008)** January, pp. 305-305.
[7] M. Cataldo, P. A. Wagstrom, J. D. Herbsleb and K. M. Carley, "Identification of coordination requirements: implications for the Design of collaboration and awareness tools", **(2006)**, pp. 353-362.
[8] M. Cataldo, J. D. Herbsleb and K. M. Carley, "Socio-technical congruence: a framework for assessing the impact of technical and work dependencies on software development productivity", Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement, **(2008)**, pp. 2-11.
[9] N. Zhou, Q. Ma and K. Ratakonda, "Quantitative modeling of communication cost for global service delivery", Services Computing, 2009. SCC '09. IEEE International Conference, **(2009)** September, pp. 388-395.
[10] E. O'. Conchuir, P. J. Agerfalk, H. H. Olsson and B. Fitzgerald, "Global software development", Communications of the ACM, vol. 52, no. 8, **(2009)** August, pp. 127.
[11] N. Nagappan, B. Murphy and V. Basili, "The influence of organizational structure on software quality", The 13th international conference, New York, USA, **(2008)**, pp. 521.
[12] J. Xu and G. Madey, "Exploration of the open source software community", Proceedings of North American Association for Computational Social and Organizational Science (NAACSOS), Pittsburgh, PA, USA, **(2004)**.

 Copyright ⓒ 2014 SERSC

Authors

Jongdae Han, received Computer Science and Engineering B.S. in 2005 and Ph.D. in 2013 from Seoul National University, Korea. He is currently an Assistant Professor at Sangmyung University, Korea. His research interests are in team composition for software development, distributed software development, and digital data mining.

Woosung Jung, received Computer Science and Engineering B.S. in 2003 and Ph.D. in 2011 from Seoul National University, Korea.

Jung was a Software Engineer at SK UBCare from 1998 to 2002 and a Senior Research Engineer in the Software Capability Development Center at LG Electronics from 2011 to 2012.

He is currently an Assistant Professor at Chungbuk National University. His research interests include software evolution, software architecture, adaptive software system, and mining software repositories.

Copyright ⓒ 2014SERSC

Copyright ⓒ 2014 SERSC