A Case Study on the Impact of Global Participation on Mailing Lists Communications of Open Source Projects

Ran Tang[1], Ahmed E. Hassan[2] and Ying Zou[1]

[1] Dept. of Elec. and Comp. Eng., Queen's University, Kingston, Ontario, Canada
{ran.tang, ying.zou}@queensu.ca
[2] School of Computing, Queen's University, Kingston, Ontario, Canada
ahmed@cs.queensu.ca

Abstract. Participants from different countries and across diverse time zones discuss important design decisions and resolve conflicts in open source projects using mailing lists. A good understanding of the social structure of these mailing lists and the impact of the global participant pool on that structure helps in managing these projects. In this paper, we present a case study which investigates the impact of global participation on communication on the developer mailing list for two large open source projects: PostgreSQL and GTK+. We find that a small group of participants from a limited number of countries dominate the mailing list while the rest of the participants contribute equally across all countries. We show that discussion threads are becoming more spread out across the globe over time. We also analyze the response delay for inquiries by newcomers to the mailing list. The delay in response to the initial inquiry depends on the country of the poster and the time when the message was posted to the list. Our findings shed light into the distribution and flow of knowledge about open source projects around the world.

Keywords: Global software development; Mining software repository

1 Introduction

Participants in an open source project heavily depend on mailing lists. These lists play a central and important role in facilitating communication among the globally distributed participants. Discussions on mailing lists often shape the future of the project and impact its progress. A good understanding of the social aspect of such discussions is needed. Prior research has demonstrated the impact of social structure on the technical structure of large software systems (e.g., [3]).

Given the global nature of open source projects, we wish to explore the social structure in the context of the global pool of the participants. In particular, we examine the interaction and communication of participants from different countries and diverse time zones on the mailing lists of open source projects. Such study helps shed light into global software development practices, and would be of great help for managers working on distributing projects across a global pool of developers [4].

In prior work [22], we developed a technique to determine the country of a mailing list participant. Using this technique we can analyze the participation and interaction patterns on mailing lists. Our technique can determine the country of 67% of participants on a mailing list. This represents an 80% improvement over prior techniques (e.g.,[8]). Using our new technique we study the impact of global participation on mailing list communications. Our findings are derived from mining a total of 20 years of the mailing list repository for two large and long-lived open source projects: The PostgreSQL and GTK+ projects. Our contributions are centered along the following three questions:

1. **What is the participation rate of countries on the mailing list?** We study whether the mailing discussions are used mostly by participants centered in a limited number of countries or distributed across the globe. We show that a small group of participants from a small number of countries dominates the list while the rest of the participants contribute equally across all countries.
2. **How global are discussions on the mailing list?** We examine if discussions show a bias to being local, i.e., given a particular discussion whether there is a tendency for participants from close by regions to participate or if the participation pattern is more global. We find that most discussion threads are well spread across the globe and that the average spread of threads increases over time. This indicates that discussions span the globe instead of being limited to specific regions.
3. **How are inquiries by newcomers handled?** We study the speed and rate of responding to inquiries by newcomers. We find that the speed and rate of responding to inquiries by newcomers depend on the country of the email sender and the time that the email was posted to the mailing list.

Organization of the paper. The rest of the paper is organized as follows. Section 2 discusses related work. Section 3 presents our research questions and the result of our case study. Section 4 discusses the threats to validity of our work. Finally, Section 5 concludes the paper.

2 Related Work

Prior studies of the global pool of participants in open source projects can be categorized into two groups: ones based on surveys and others based on mining repositories.

Studies based on surveys. Robles et al. [14] surveyed over 5,500 respondents to identify their country. Robles et al. showed that a majority of open source developers are from Europe. Similar results are also reported in the survey conducted by Ghosh [7] and David et al. [5].

Studies based on mining repositories. Dempsey et al. [6] analyzed the top-level domain name of the email address (e.g., .ca, .com) of a participant to identify the country of the participant. However the study did not compensate for the US bias resulting from the wide use of generic domains (e.g., .com). Prior techniques do not map .com addresses to any country. Therefore, the participants from the US may be under represented in the analysis. Studies in [12, 14] show that the developer pool is becoming more European-based over time. Robles and Gonzalez-Barahona [8, 13] used a technique to identify countries of participants in SourceForge [21] open source projects. Email address and time zone information in the user profile are analyzed to infer the country. The mailing list was also studied using a similar technique. However, the time zone information in the mailing list does not contain specific country information. Therefore, the analysis of the time zone can only derive the origins of participants to specific time zone regions instead of particular countries.

Other work on mining mailing list. Several studies mine mailing list repositories. For example, Mockus et al. [16] conduct two case studies to reveal the process of open source development using mailing list repository. Bird et al. [1] build social networks using information derived from the PostgreSQL mailing list. These studies do not explore the impact of geographical distribution on the social interaction of participants.

3 Case Study

We conducted a case study to explore various aspects of participation and interaction in the mailing lists. We use the developer mailing lists for the PostgreSQL [20] (postgresql-hackers) and GTK+ [9] (gtk-devel-list) projects in our case study.

Table 1. Statistics about the studied mailing lists

	Studied Period	**# of Participants**	**# of Threads**
PostgreSQL	1999-2008	4,742	23,104
GTK+	1999-2008	2,734	7,481

Table 2. Research questions

	Research Questions
Q1	What is the participation rate of countries on the mailing list?
Q2	How global are discussions on the mailing list?
Q3	How are inquiries by newcomers handled?

The PostgreSQL project is a relational database management system. The GTK+ project is a toolkit for creating cross platform graphical user interfaces. Both projects involve a large pool of international developers who interact through the mailing lists. Both projects come from two different domains: database management and graphic user interface development. Our objective is to study if our results hold across domains and projects. Table 1 presents descriptive statistics about both projects. Using the recovered countries for the participants, we sought to explore the research questions listed in Table 2. For each question, we present our motivation, and discuss our results using data from the mailing list repositories in the PostgreSQL and GTK+ projects.

4.1 What is the participation rate of countries on the mailing list?

Motivation. Prior research shows that open source projects have a small core team and a small number of core contributors [16, 19]. We want to identify the core participants in the mailing lists and study their distribution around the world. We wish to compare the participation of that small core with the rest of the participants. In particular, we want to examine if they are localized to small number of countries or if they are distributed around the globe. Such knowledge would be helpful in the planning and recruiting processes for open source projects. For example, the knowledge of global distribution and involvement of participants may help conference planner select an optimal locations for face-to-face project conferences (e.g.: PostgreSQL Conference [18]) in order to achieve high attendance.

Results. We measure the number of participants who contribute the majority (i.e., 70%) of the messages to the mailing list. We call these participants the core participants (similar to [16]). Table 3 shows that although these core participants represent a small percentage (i.e., 1.5-5%) of all the participants, they are spread out over a relatively larger percentage of countries.

Table 3. Statistics of core participants

	#participants (%participants)	**# countries (%countries)**
PostgreSQL	47 (1.5%)	13 (13%)
GTK+	96 (5%)	21 (27%)

Table 4 and Table 5 show in more detail the countries of all participants. We observe that the participation patterns vary between projects. For instance, although the US has the highest number of participants in both projects, the number of messages sent by US participants varies considerably in both projects. While the US participants contributing most (~58%) of the messages on the PostgreSQL mailing list, they only contribute ~20% of the messages on the GTK+ mailing list with Germany being the top contributor of messages.

Table 6 and Table 7 demonstrate an interesting pattern for the contribution of countries to the mailing list. Looking at the median of the number of messages and threads in each country, we find that the median is surprisingly very low and that it is consistent across countries. The majority of participants post 1 or 2 messages and are involved in 1 or 2 threads. This pattern leads us to hypothesize that most participants

Table 4. Country composition of the PostgreSQL mailing list

Country	Participants (%)	Msgs (%)
United States	1037(32.6%)	76723(57.8%)
Germany	228(7.2%)	7237(5.5%)
Canada	160(5.0%)	7602(5.7%)
UK	144(4.5%)	8584(6.5%)
Australia	108(3.4%)	4862(3.7%)
Russia	98(3.1%)	2578(1.9%)
India	97(3.0%)	574(0.4%)
France	97(3.0%)	1621(1.2%)
Italy	92(2.9%)	424(0.3%)
Brazil	90(2.8%)	424(0.3%)
Japan	89(2.8%)	3979(3.0%)
Netherlands	66(2.1%)	722(0.5%)
China	54(1.7%)	210(0.2%)
Poland	51(1.6%)	326(0.2%)
Czech	48(1.5%)	940(0.7%)
Austria	47(1.5%)	3247(2.5%)
Sweden	44(1.4%)	2974(2.2%)
Hungary	41(1.3%)	271(0.2%)
Spain	37(1.2%)	227(0.2%)
Denmark	28(0.9%)	209(0.2%)
New Zealand	28(0.9%)	1,024(0.8%)
Other	492(15.5%)	7,891(6.0%)

Table 5. Country composition of the GTK+ mailing list

Country	Participants (%)	Msgs (%)
United States	517(27.8%)	4,623(19.9%)
Germany	189(10.2%)	6,670(28.7%)
France	124(6.7%)	1026(4.4%)
UK	120(6.5%)	3,111(13.4%)
Sweden	64(3.4%)	800(3.4%)
Australia	63(3.4%)	708(3.0%)
Canada	57(3.1%)	429(1.8%)
Italy	55(3.0%)	260(1.1%)
India	53(2.9%)	173(0.7%)
Netherlands	50(2.7%)	268(1.1%)
Spain	42(2.3%)	216(0.9%)
China	41(2.2%)	1469(6.3%)
Finland	32(1.7%)	864(3.7%)
Russia	29(1.6%)	242(1.0%)
Brazil	27(1.5%)	147(0.6%)
Japan	25(1.3%)	110(0.5%)
Austria	23(1.2%)	65(0.3%)
Belgium	23(1.2%)	134(0.6%)
Czech	22(1.2%)	113(0.5%)
Norway	21(1.1%)	99(0.4%)
Other	283(15.2%)	1,733(7.5%)

Table 6. Participation level for the PostgreSQL mailing list

Country	Mean Msgs	Med Msgs	Mean Threads	Med Threads
US	74	2	32.9	1
Germany	31.7	2	17.3	1
Canada	47.5	2	24.3	2
UK	59.6	3	27.6	2
Australia	45	2	24.8	1
Russia	26.3	2	14.4	1
India	5.9	1	3.5	1
France	16.7	2	8.7	1
Italy	4.6	2	3	1
Brazil	4.7	2	3.1	1
Japan	44.7	2	19.	2
Netherlands	10.9	2	6.3	2
China	3.9	2	3	1
Poland	6.4	2	3.4	1
Czech	19.6	3	11.1	2
Austria	69.1	4	42.7	2
Sweden	67.6	2.5	30.1	1.5
Hungary	6.6	2	3.4	1
Spain	6.1	1	3.3	1
Denmark	7.5	3.5	5.2	2
New Zealand	36.6	2.5	19.2	1.5

Table 7. Participation level for the GTK+ mailing list

Country	Mean Msgs	Med Msgs	Mean Threads	Med Threads
US	8.9	2	5.8	2
Germany	35.3	2	21.7	2
France	8.3	2	5.6	1
UK	25.9	2	15.2	2
Sweden	12.5	2.5	8.4	2
Australia	11.2	2	8.5	1
Canada	7.5	2	4.3	1
Italy	4.7	2	2.8	1
India	3.3	1	2.7	1
Netherlands	5.4	1	3.6	1
Spain	5.1	2	3.6	1
China	35.8	1	21.3	1
Finland	27	2.5	16.6	1.5
Russia	8.3	2	4.8	1
Brazil	5.4	2	3.1	2
Japan	4.4	2	2.5	1
Austria	2.8	1	2.1	1
Belgium	5.8	2	3.6	1
Czech	5.1	2	3.6	1
Norway	4.7	1	2.3	1

rarely use the mailing list for discussion. Instead they post 1 or 2 inquiries in the mailing list. In short, most mailing list participants use it to post inquiries, rather than to delve into in-depth discussions.

These observations about the different use of mailing lists by core members and newcomers shape our next two questions. Q2 will study the spread of countries in in-depth discussion threads while Q3 will study the speed of response to inquiries by newcomers.

Developer mailing lists are dominated by a very small number of participants who are from a relatively larger number of countries. Participants contribute equally to the mailing lists independent of their country.

4.2 How global are discussions on the mailing list?

Motivation. The results for our previous question indicate that the participant pools in both projects are globally distributed. However, how these participants interact remains unanswered. Do participants talk globally or do they prefer to talk to participants locally? This question examines the interaction in the global open source development. This helps us gain insight about the problems of coordination and localization of knowledge for distributed teams.

Results. We define a spread metric for each thread to measure the global spread of a thread. A discussion thread is a collection of email messages related to each other by replying. A participant starts a discussion thread by posting a question or raising an issue, and other participants may choose to reply to it. By examining the diversity of the participants' locations in each thread, we can determine whether the discussion is global or primarily localized.

We use the MESSAGE-ID field in an email to reconstruct discussion threads since each message, as part of a thread, would refer to the message id of an earlier message in the thread. Sometimes, a thread is re-opened for some reason. For example, participants may reply to a thread which has had no postings for more than one year. Since the discussion has stopped for too long, this reply essentially creates a new discussion on the same topic. We process such a reply as the starting point of a new thread if the time between the reply and the last posting is a long period of time. We use a threshold of 30 days to cut off re-opened threads into two different threads. This threshold is selected by manually examining the re-opened threads in both studied projects.

To compute the spread of a thread, we calculate the spread between each pair of participants who have posted on the thread. As shown in Figure 1, for each pair of participants A and B, we use either the time zone difference from A to B (clockwise) or B to A (clockwise), whichever is less, as the spread between them. The maximum spread between two participants is 12. The spread is 0 if both participants are in the same time zone. In the example shown in Figure 1, the spread is 8. We pick the largest spread between all pairs of participants in a thread. A large spread is a good indicator of the global spread of interaction in a thread.

We consider threads with a spread less or equal to 5 as low spread threads. Typically, such threads (i.e., with the spread of 5) represent discussion within one country (e.g., US) or one close region (e.g., EU). We then examine the trend of low spread threads relative to all threads over time.

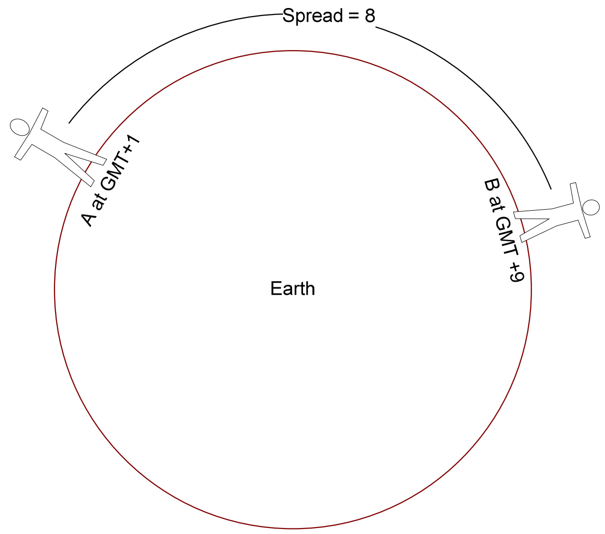

Figure 1: An example of spread calculation

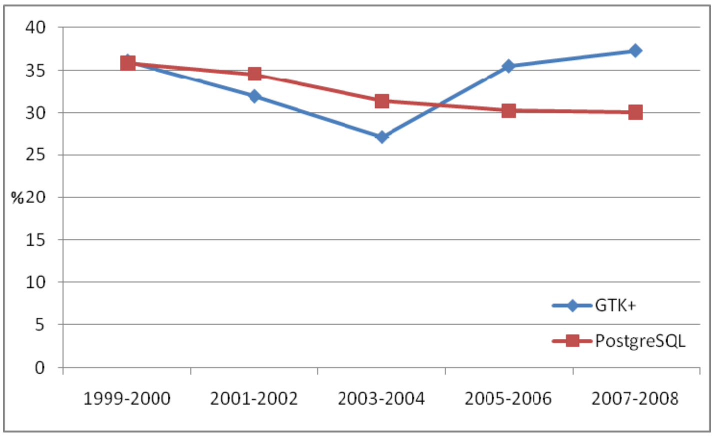

Figure 2: % of low-spread threads over time

As shown in Figure 2, low spread threads only represent about one third of the total threads and low spread threads decrease in the PostgreSQL project over time. The participants of PostgreSQL list were primarily from the US when it was started in

1999. The mailing list attracts more participants throughout the world and the PostgreSQL project is becoming more international over time.

The GTK+ project illustrates another interesting trend. In early years, it follows a similar trend with the project being localized in the US. We then note that the percentage of low-spread threads continues to decrease. This project was at first centralized in the US, and then it grew to have European participants till 2003. However, this trend is reversed with more localized European threads since 2003. An observation about the project population of GTK+ and other GNOME projects was noted by [12] which mined the source code change logs of the credit, instead of examining the mailing lists participants. The confirmation of our mailing list findings by mining the source code repositories demonstrates the importance of social information in explaining and collaborating information recovered from other project repositories.

Discussion threads tend to become more global as a project evolves. However, this trend might change over time for some projects.

4.3 How are inquiries by newcomers handled?

Motivation. Prior research [13] and our earlier results show that many open source projects have a high concentration of participants from US, Canada or European Union (EU). We believe that the high concentration of participants might impact the openness of such projects. The openness of the mailing list is an important factor that influences newcomers. For example, if inquiries by newcomers are often ignored or take a long time to get a response, then newcomers might lose their interest in the project and not join the community [2]. Two factors which might affect the delay in responding to an initial inquiry are the country of the poster and the posting time (which might be indirectly affected by the country). For example, posts by Chinese participants might not get an immediate response till next day when the North American participants are at work.

Table 8. Response delay and ratio for both studied projects

	EU, US, Canada		Others	
	Response Delay (hours)	**Response Ratio (%)**	**Response Delay (hours)**	**Response Ratio (%)**
PostgreSQL	0.17	68.2%	1.66	67.0%
GTK+	0.46	61.9%	19.7	59.0%

Results. In our analysis, we divide all countries into two groups:

1. Group one includes countries from the EU, the US and Canada. This group includes many developed countries which are known to be active in open source development [8].

2. Group two includes all other countries. This group mostly includes developing countries with less open source development activity.

We measure the median response delay and the response ratio for each group. We define the response delay for an initial inquiry as the time difference between the initial inquiry and the first reply. The response ratio is denoted as the number of replied inquiries divided by the total number of inquiries throughout the studied period for each project. We only examine participants who have less than 20 messages in the mailing list. We choose not to consider core participants with more messages since we believe that the rest of the mailing list is familiar with them and that they would receive a response independent of their country or time of posting.

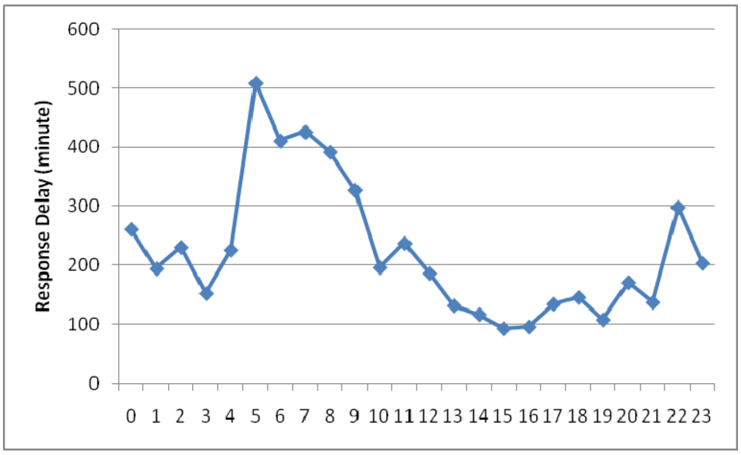

Figure 3. Response delay for the PostgreSQL project (GMT)

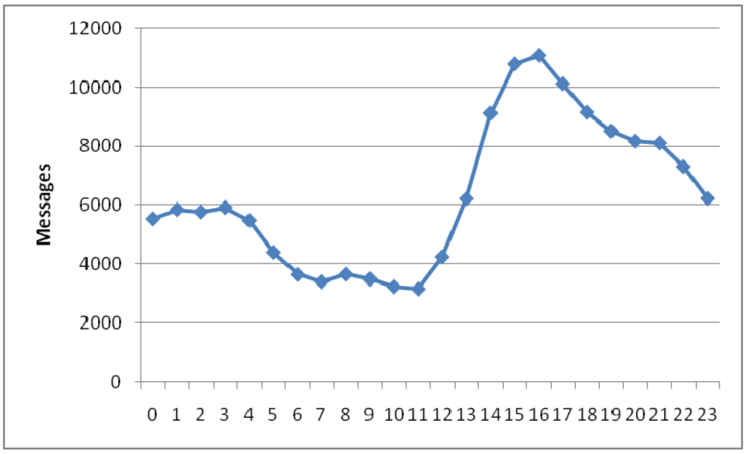

Figure 4. Traffic on the PostgreSQL mailing list (GMT)

As shown in Table 8, newcomers from the EU, the US and Canada get a statistically significant faster response compared to those from other countries. As for the response ratio, the ratios are consistent across both groups with very low variations that are not statistically significant. The response delay and ratio for the PostgreSQL project are much lower than the GTK+ project. One possible reason is the fact that the PostgreSQL project has a larger participant pool which is more distributed as shown in Q2.

We plotted the response delay for both projects throughout a day. Figure 3 depicts the response delay for the newcomers in the PostgreSQL project. There is a large spike in delay between 5 to 11 GMT. In an effort to investigate this spike in delay, we plotted the traffic (i.e., the total number of messages posted per GMT hour of a day) in the PostgreSQL mailing list shown in Figure 4. The patterns in Figure 3 and Figure 4 are inverted. To better quantify the relation, we calculated the spearman correlation between both metrics plotted in Figure 3 and Figure 4. The correlation is -0.79 for the PostgreSQL project and -0.59 for the GTK+ project, indicating a strong negative correlation. In short, when the traffic is high, the response delay is low; and vice versa. Although one might assume that an open source mailing list provides around the clock support, the mailing list in many ways operates as a traditional company which has specific support hours and reduces staff in the off-peak hours. For most participants, it is of little value to post a message in off-hours, since there is a high chance that the message won't receive a reply till the list is active again (probably the following day).

Our analysis shows that the response delay depends on the country and the posting time over the participants from all countries. We sought to explore the response delay in a particular country. We picked the US with a sufficient number of inquiries that can be spread over 24 hours. We studied its response delay pattern. As depicted in Figure 5, the response delay pattern is similar to the pattern described in the overall graph (Figure 3) for the PostgreSQL project. Therefore, we believe that the response delay for an active country also depends on the posting time.

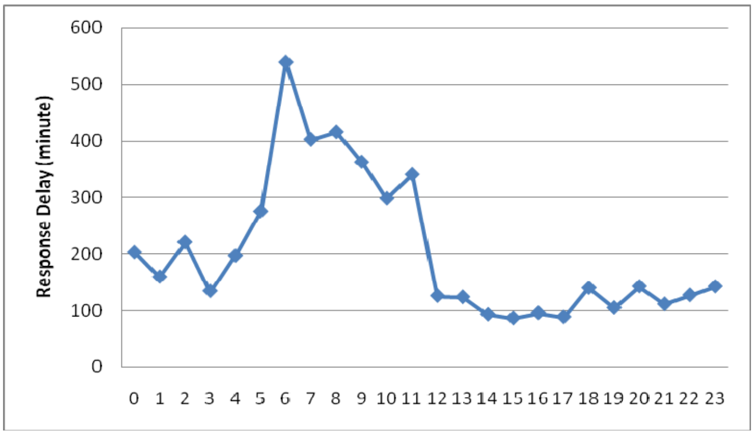

Figure 5. Response delay over one day for US based participants for the PostgreSQL

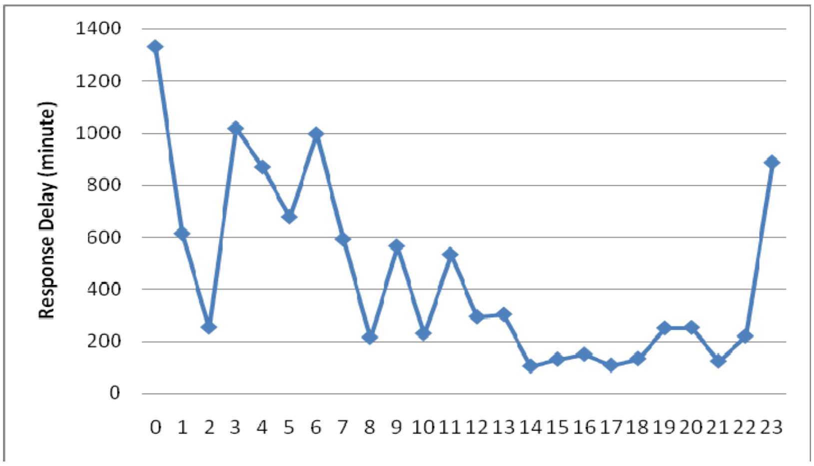

Figure 6. Response delay over one day for US based participants for GTK+

> The response delay and ratio to a newcomer's initial inquiry depend on the time when the inquiry was posted and on the country of the participant who posted it. Open source mailing lists do not operate at full capacity (e.g., 24 hours over 7 days). Instead the open source mailing lists operate in the similar way as support lines with high and low staffing periods based on the time of a day.

5 Threats to validity

Our work has several limitations which affect the validity and the generality of our findings. First, our findings are based on studying two open source projects. We chose two long-lived successful and active open source projects for our study with a well-archived mailing list repository. In future work, we need to explore additional projects to verify the generality of our findings. It would be interesting to explore non-successful projects though the mailing lists of such projects are not as active and likely won't have as much discussions.

Our approach [22] to identify the location of a participant use IP2Location databases [11]. These databases are built using several heuristics and might contain errors [17]. Moreover, it might be the case that the location of a specific IP has changed over time with the IP2Location database mapping an out-of-dated IP address to the most recent location recorded in the database. Multi-national companies might have their whole intranet accessing the internet through US-based gateways. This would cause all remote offices to appear as if they are in the US. We determine the location of a participant based on the most frequently reported country using the sender IP address analysis. However, a participant may move from one country to another over the years. We also assume that each participant can reside in a single country. A cursory analysis of the data shows that almost all participants have 90% of their posts coming from the same country. Nevertheless, we plan to explore these assumptions in future work.

A final limitation of our approach is that the correct country of a participant is not known. In other words, there exists no gold standard to compare against. As a basic accuracy verification of our approach, we compared the identified participants by both our approach and prior work [8]. We found that both approaches have a 4-5% mismatch ratio in identified countries. Studying the mismatches, we find that 70-80% of them are due to our approach using the most frequent IP location of the sender instead of mapping a participant to the country indicated by resalable country domain name.

We have defined a few thresholds, such as the low spread threads (less than 5 time zones) and the intervals for creating a new thread from an inactive thread (i.e., 30 days). These thresholds work well in analyzing the two studied projects. In the future, we plan to examine other possible thresholds on more projects.

Much of our findings show correlation between attributes without explaining the causes. More studies are needed to explore the causes. For example, our findings show that the initial inquiries submitted by the participants from the regions outside of the US, Canada, and the EU have a high response delay. However, we need to further explore the reasons through ethnographical studies.

6 Conclusion

Studying communication on mailing lists shed light into the spread and flow of knowledge for a project. Through a case study on two large and long-lived open source projects: PostgreSQL and GTK+, we investigated the impact of having globally distributed participants communicating on the mailing list. We found that a small number of participants spread over a larger set of countries dominate the discussion. We found that the majority of participants contribute a single message to a single thread. We noted and examined two different uses of the mailing list: a) for lengthy discussions by core members; b) for inquiries by newcomers. For lengthy discussions, we found that over time the discussions become more spread out across the globe for one of the studied projects, while the other project (GTK+) has the same trend in the beginning but later becomes less spread out. A closer analysis indicates that mailing list participants often reflect the developer composition of a project. In the case of the GTK+ project, our analysis noted the migration of the development team from the US to the EU. As for inquiries by newcomers, we found that delays in responding to such inquiries depend on the country of the newcomer and the posting time of the inquiry. Our results help us better understand the social structure and global nature of open source projects, and their impact on timely and open discussions in open source projects.

References

1. Bird, C., Gourley, A., Devanbu, P., Gertz, M., and Swaminathan, A. Mining email social networks in Postgres. International Workshop on Mining Software Repositories, 2006.
2. Bird, C., Gourley, A., Devanbu, P., Swaminathan, A., Hsu, G. Open Borders? Immigration in Open Source Projects, International Workshop on Mining Software Repositories, 2007.
3. Bowman, I. T. and Holt, R. C. Software architecture recovery using Conway's law. Conference of the Centre For Advanced Studies on Collaborative Research, 1998.
4. Cherry, S., Robillard, P. N. Communication Problems in Global Software Development: Spotlight on a New Field of Investigation. International Workshop on Global Software Development, 2004.
5. David, P.A., Waterman, A., Arora, S. FLOSS-US. The free/libre/open source software survey for 2003. Technical Report, Stanford Institute for Economic and Policy Research, 2003.
6. Dempsey, B.J., Weiss, D., Jones, P., Greenberg, J. Who is an open source software developer? Communications of the ACM, 45(2), 2002.
7. Ghosh, R.A., Glott, R., Krieger, B., Robles, G. Survey of developers (free/libre and open source software: survey and study). Technical Report, University of Maastricht, 2002.
8. Gonzalez-Barahona, J. M., Robles, G., Andradas-Izquierdo, R. and Ghosh, R. A., Geographic origin of libre software participants, Information Economics and Policy, 20 (4), December 2008.
9. GTK+ developer mailing list, http://mail.gnome.org/mailman/listinfo/gtk-devel-list/, last accessed on October 29 2009.
10. Hertel, G., Niedner, S., Herrmann, S. Motivation of software developers in open source projects: an internet-based survey of contributors to the linux kernel. Research Policy, Volume 32, 2003.
11. IP2Location Database, http://www.ip2location.com/, last accessed on October 29 2009.
12. Lancashire, D. Code, culture and cash: the fading altruism of open source development. First Monday, 6 (12), December 2001.
13. Robles, G., Gonzalez-Barahona, J. M. Geographic location of developers at SourceForge. International Workshop on Mining Software Repositories, 2006, pp. 144-150.
14. Robles, G., Scheider, H., Tretkowski, I., Weber, N. Who is doing it? A research on libre software developers. Technical Report, Technische Universitt, 2001.
15. Member states of the EU, http://europa.eu/abc/european_countries/index_en.htm, last accessed on October 29 2009.
16. Mockus, A., Two case studies of open source software development: Apache and Mozilla, ACM Transactions on Software Engineering and Methodology, 11(3), Jul. 2002.
17. Padmanabhan, V., Subramanian, L. An investigation of geographic mapping techniques for Internet hosts. In Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, 2001.
18. PostgreSQL conference, http://postgresqlconference.org/ last accessed on October 29 2009.
19. PostgreSQL contributor profiles, http://www.postgresql.org/community/contributors/, last accessed on October 29 2009.
20. PostgreSQL developer mailing list, http://archives.postgresql.org/pgsql-hackers/, last accessed on October 29 2009.
21. SourceForge open source project repository, http://sourceforge.net, last accessed on October 29 2009.
22. Tang, R., Hassan, A.E., Zou, Y., Techniques for Identifying the Country Origin of Mailing List Participants, 16th Working Conference on Reverse Engineering, 2009
23. Tuomi, I. Evolution of the Linux credits file: methodological challenges and reference data for open source research. First Monday, 9(6), June 2004.