Analyzing OSS Developers' Working Time Using Mailing Lists Archives

Masateru Tsunoda
Nara Institute of Science and Technology
Kansai Science City, 630-0192 Japan
masate-t@is.naist.jp

Akito Monden
Nara Institute of Science and Technology
Kansai Science City, 630-0192 Japan
akito-m@is.naist.jp

Takeshi Kakimoto
Nara Institute of Science and Technology
Kansai Science City, 630-0192 Japan
takesi-k@is.naist.jp

Yasutaka Kamei
Nara Institute of Science and Technology
Kansai Science City, 630-0192 Japan
yasuta-k@is.naist.jp

Ken-ichi Matsumoto
Nara Institute of Science and Technology
Kansai Science City, 630-0192 Japan
matumoto@is.naist.jp

Categories and Subject Descriptors
K.6.1 [**Management of Computing and Information Systems**]: Project and People Management – *Staffing;*

General Terms: Management

Keywords
Overtime work, workload

1. INTRODUCTION
We chose PostgreSQL, a relational database system for the MSR mining challenge.

Our research question is in the following mining area:

- Process analysis

Our mining question is "when OSS developers work?" OSS developers' working time may be a good indicator to understand the development style of a project. (For example, if many developers work in office hour, these might be daily works in a company.)

2. INPUT DATA
We used mailing lists (MLs) archives of PostgreSQL, downloaded from http://www.postgresql.org/community/lists/. The MLs mainly consist of user lists and developer lists. We used developer lists archive since we needed developers' working time. Table 1 explains details of each ML. Figure 1 shows amounts of messages of each ML in the developer lists. Amounts of messages were increasing year by year. The ML of hackers had many more messages than other MLs. We extracted MLs archives till December 2005. Note that most of committers' messages were automatically generated when source code was checked into software configuration management repository.

MSR'06, May 22–23, 2006, Shanghai, China.
ACM 1-59593-085-X/06/0005.

Table 1. Description of Each ML

List	Description	Archived from
committers	Notification of CVS commits are sent to this list.	April 2000
hackers	Discussion of current development issues, problems and bugs, and proposed new features.	January 1997
patches	Patches for new features and bug fixes should be sent to this list.	June 2000
www	Discussion of development and coordination of the PostgreSQL websites.	August 2003

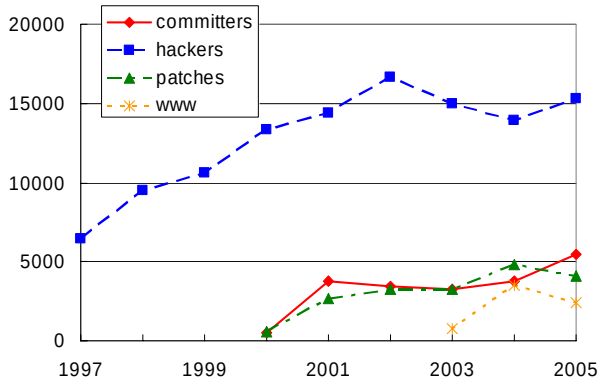

Figure 1. Amount of Messages in Each Year

We picked up "mail sent time" to identify developers' working time. Getting mail sent time from the MLs archives consists of the following two steps: First, we downloaded the MLs archives with

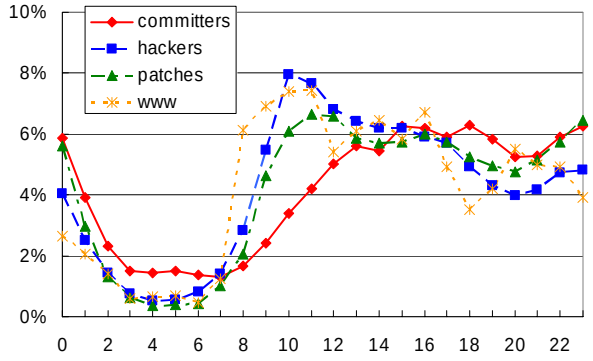

Figure 2. Ratio of Messages in Each Hour

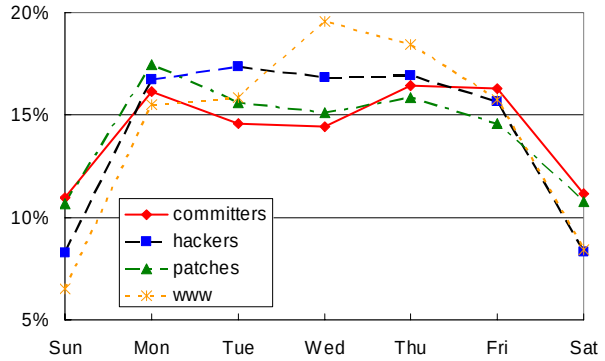

Figure 3. Ratio of Messages in Each Day of a Week

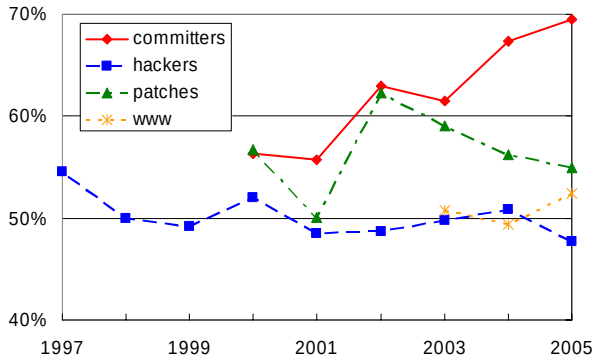

Figure 4. Ratio of Messages Sent at Overtime Period in Each Year

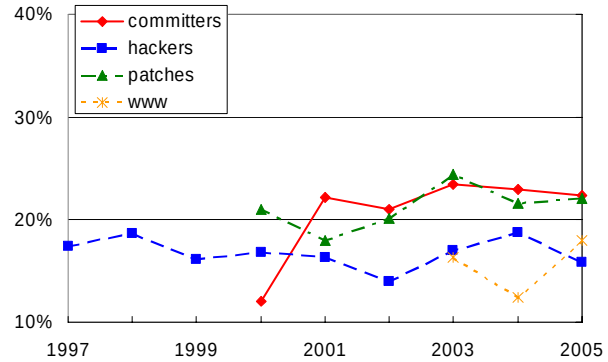

Figure 5. Ratio of Messages Sent on Weekend in Each Year

Irvine[1], a web download tool. Then, we extracted mail sent time from the downloaded archives with a Perl script.

In our analysis, we mainly focused on the following aspects:

- Mail sent hour
- Days of a week of mail sent date
- Difference in hours and date among ML groups
- Time trend

To see developer workload, we defined the overtime period. Overtime period includes before 9a.m. and after 5p.m. on weekday, and all day of weekend. Because each ML has different amount of messages, we used ratio of messages, defined as amount of messages divided by total amount, of each ML group.

3. RESULTS AND INTERPRETATIONS

The ratio of messages in each hour is shown in Figure 2. The ML of hackers was active in the morning. On the contrary, ratio of committers' messages in the evening is comparatively higher than other ML groups. The ratio of messages in each day of a week is shown in Figure 3. Most developers work on weekday. The ratio of committers' messages on weekend is slightly higher than other ML groups.

Ratio of messages sent at overtime period in each year is shown in Figure 4. The ratio of messages of committers was increasing year by year while other ML groups did not show clear trends. This may suggest that many committers are recently required to work at overtime period by some reasons (e.g. too many patches to inspect). Ratio of messages sent on weekend by year is shown in Figure 5. In spite of the increase of committers' messages sent at overtime period, committers' messages sent on weekend did not increase. This may suggest that even if committers are willing to work at overtime period, they do not want to work on weekend.

4. CONCLUSIONS

We analyzed mailing lists archives of PostgreSQL. We focused on mail sent hour, days of a week of mail sent date, difference in hours and data among ML groups, and time trend. Our finding is that the ratio of committers' messages sent at overtime period was increasing year by year.

5. ACKNOWLEDGMENTS

This work is supported by the EASE (Empirical Approach to Software Engineering) project of the Comprehensive Development of e-Society Foundation Software program of the Ministry of Education, Culture, Sports, Science and Technology of Japan.

[1] http://hp.vector.co.jp/authors/VA024591/ (in Japanese)