Temporal Changes of the Openness of an OSS community: A Case Study of the Apache HTTP Server Community

Mizuki Yamamoto, Masao Ohira, Yasutaka Kamei
Shinsuke Matsumoto and Ken-ichi Matsumoto
Graduate School of Information Science Nara Institute of Science and Technology
{ mizuki-y, masao, yasuta-k, shinsuke-m, matumoto } @ is.naist.jp

1. Introduction

The development style of open source software (OSS) is widely recognized as "bazaar-like" software development [1]. In "bazaar-like" software development, there is openness to anyone who wants to participate in a community. Participants including not only developers but also users share, discuss, and implement a variety of ideas with each other, and then release OSS products to the public. Certain products (e.g. Linux, Apache) developed by "bazaar-like" software development are widely used in our daily life. However, at the same time, a larger number of OSS products and communities following the "bazaar-like" style result in failure or stagnation due to various factors [2].

This implies that "bazaar" style of software development might not be a critical success factor of OSS products and communities. If so, how could certain products and communities achieve great success and survive in a free competitive market for more than a decade? Our study is motivated to answer this question.

2. Hypothesis

In order to answer the research question, we develop the basic hypothesis that *the condition of "bazaar" (i.e., the openness of "bazaar") in a successful OSS community changes over time in evolving to adapt itself to the external environment.* We consider that the openness of "bazaar" is determined by ease of community participation and acceptability of contributions from non-core members. The ease of community participation would be ensured because anybody can subscribe to mailing lists and send a patch. By contrast, the acceptability of contributions from non-core members might not be guaranteed. Therefore, we analyze acceptability of contributions from non-core members and develop more concrete hypotheses as follows.

[H1] *A community with a (small/big) difference in the acceptability between core and non-core members[1] will (acquire/lose) non-core members over time.*

[H2] *A community with a (small/big) difference in the acceptability between core and non-core members will (increase/decrease) in number of contributions from non-core members over time.*

3. Case study and Results

Our case study targeted the Apache HTTP Server community which has been developing web server software with the biggest market share. We analyzed the developer's mailing list for 11years from Jan. 1, 1997 to Dec. 31, 2007.

3.1. Number of developers

Figure 1 shows changes in the number of core developers and non-core developers. The x-axis shows time and the y-axis shows the number of developers. The number of non-core developers rapidly increased until 2002 and decreased until 2007. On the other hand, the number of core developers increased steadily (15 to 50 developers). This result can be seen as the total growth of the Apache community.

3.2. Number of patches

Figure 2 shows the number of patches sent by core developers and non-core developers. The x-axis shows time and the y-axis shows the number of patches. The total number of patches during the analysis period was 3061. The core developers sent 1820 (about 60%) of the total number of patches.

In the early periods (1997-1998), most patches (about 86%) were sent by about 20 core developers. This result indicates most defects were detected and corrected by the core developers in the early periods. In the middle periods (1999-2002), non-core developers began to send many patches, while the number of patches sent by core developers significantly decreased. The total number of patches has decreased since 2002, and eventually only 52 were sent in 2007.

3.3. Review rate

Figure 3 shows the review rates (i.e., percentage of patches that were reviewed by other developers) of core developers and non-core developers. The values written at the top of Figure 3 show the p-value calculated by

[1] Core members have a write access right to the code repository but non-core members don't have.

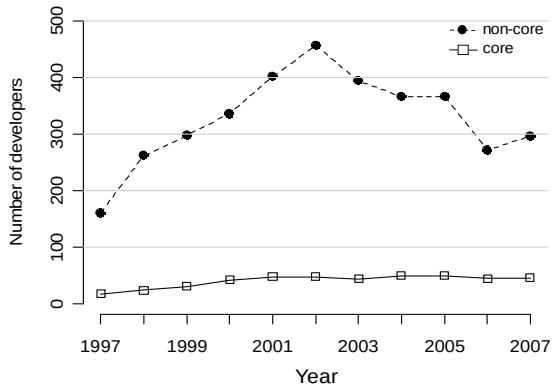

Figure 1. Changes in number of developers

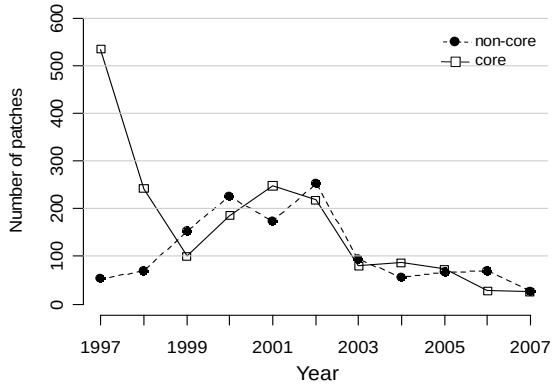

Figure 2. Changes in number of patches

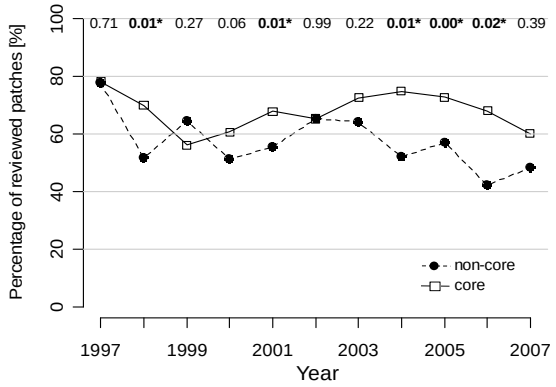

Figure 3. Changes in percentage of reviewed patches

chi-square test. Bold letters indicate statistically significant ($P < 0.05$) differences between the review rate of core developers and non-core developers.

In 1997, the review ratio of core developers was notably higher (80.0%) than other periods, although a large number of patches (over 500) were sent by core developers. In the later periods (2002-2007), there were significant differences between the review rate of core and non-core developers. The review ratio of non-core was significantly low from 2004 to 2006. In particular, over the half of the patches sent by non-core developers had received no response in 2006. As a result, the review rate of non-core developer was the lowest in 2006.

3.4. Hypotheses testing

We tested our two hypotheses described in section 2. In Figure 3, the differences in the review ratio between core and non-core developers are relatively small until around 2002 and become large from 2003. While Figure 1 shows the rapid increase of non-core developers until 2002, it also showed the decrease of non-core developers. We can consider that the hypothesis [H1] is supported. In Figure 2, the number of patches from non-core developers increased until 2002 and decreased from 2002. As with [H1], we can conclude that the hypothesis [H2] is supported.

4. Discussions and Future work

In the first half of the analysis period (1997-2002), the difference of the acceptability between core and non-core developers in the Apache community can be said as relatively small. The result suggests that there is a positive stance toward widely taking contributions not only from core developers but also from non-core developers in this period, that is, the condition of the community was open. In the last half of the analysis period (2003-2007), however, the difference of the acceptability is relatively big. The difference of the review rate could be confirmed three times during 5 years. We consider that the community in this period lost the positive attitude as described above and the openness of the community also decreased. These factors would strongly affect the decreasing number of non-core developers and patches from non-core developers.

Summarizing these results, we can conclude that the Apache community changes own openness over time to control the condition of the community, in other words, to survive in the rapidly changing external environment.

In the near future, we need to analyze not only the Apache community but also other OSS communities to assure the generality of the conclusions.

Acknowledgement

This research is being conducted as a part of the Next Generation IT Program, Grant-in-aid for Young Scientists (B), 20700028, 2008 and Grant-in-Aid for Japan Society for the Promotion of Science (JSPS) Fellows (Research No:20009220) by the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

[1] E.S. Raymond, The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary, O'Reilly and Associates, 1999.

[2] M. Ohira, N. Ohsugi, T. Ohoka and K. Matsumoto, "Accelerating cross-project knowledge collaboration using collaborative filtering and social networks", *Proceedings of International Workshop on Mining Software Repositories (MSR2005)*, St.Louis, MO., pp.111-115, 2005.