

Poster: Companies' Domination in FLOSS Development – An Empirical Study of OpenStack

Yuxia Zhang, Xin Tan, Minghui Zhou, Zhi Jin

Key Laboratory of High Confidence Software Technologies, Ministry of Education
Institute of Software, EECS, Peking University, Beijing, China
{yuxiaz,tanxin16,zhmmh,zhijin}@pku.edu.cn

ACM Reference Format:

Yuxia Zhang, Xin Tan, Minghui Zhou, Zhi Jin. 2018. Poster: Companies' Domination in FLOSS Development – An Empirical Study of OpenStack. In *ICSE '18 Companion: 40th International Conference on Software Engineering Companion*, May 27–June 3, 2018, Gothenburg, Sweden. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3183440.3195047>

1 INTRODUCTION

Because of the increasing acceptance and possibly expanding market of free/libre open source software (FLOSS), the spectrum and scale of companies that participate in FLOSS development have substantially expanded in recent years. Companies get involved in FLOSS projects to acquire user innovations [3, 12], to reduce costs [8, 11], to make money on complementary services [13], etc. Such intense involvement may change the nature of FLOSS development and pose critical challenges for the sustainability of the projects. For example, it has been found that a company's full control and intense involvement is associated with a decrease of volunteer inflow [13]. Sometimes a project may fail after one company pulls resources from the project [13]. This raises concerns about the domination of one company in a project. In large projects like OpenStack, there are often hundreds of companies involved in contributing code. Despite substantial researches on commercial participation, whether or not one company dominates a project and the impact of such domination has never been explicitly explored. We investigate four main projects of OpenStack, a large ecosystem that has had a tremendous impact on computing and society, to answer the following research questions: Does one company dominate the project's development (RQ1)? If the answer to RQ1 is yes, does the domination affect the community (RQ2)?

2 METHODOLOGY

OpenStack is a cloud computing infrastructure capable of handling big data [7]. We conduct an empirical study on four representative projects of OpenStack (Nova, Swift, Glance, and Neutron) to explore the commercial domination. A company's investment in a project involves many resources, but the code and manpower are the most critical elements in the FLOSS development [13]. Thus, we employ the development data of the four projects to quantify companies' contributions from two aspects: the developers they provided and the commits made by those developers. Due to space limitation,

we only introduce the metrics and results concerning the aspect of commits, while the aspect of developers presents the similar results.

2.1 Data Preparation

The dataset is obtained from an official website¹ maintained by the OpenStack community containing two parts: commits and issue reports. We take several steps to clean and standardize the raw data for further analysis, including: 1. dropping the commits or issue reports submitted by aided OSS development tools, like Gerrit²; 2. merging multiple identities of one contributor by using a semi-automatic process [2]; 3. identifying the developers' affiliated companies by using a member list³ maintained by the OpenStack community.

2.2 Metrics

To answer RQ1 and RQ2, we need to construct metrics to characterize one company's domination in a FLOSS project by borrowing the ideas of "code ownership" and "Shannon entropy" [1, 4, 9].

1. Domination Intensity. This metric characterizes the degree of contribution made by the company with the most contributions. Based on the concept of code ownership, we define the domination intensity of commits as below:

$$CommitsIntensity^p = \max \left(\frac{num_of_commits_c^p}{total_num_of_commits^p} \right)$$

where $CommitsIntensity^p$ is the contribution ratio of the number of commits from company c to the number of total commits of the project p .

2. Domination Entropy. Based on the concept of Shannon entropy, we design a metric to describe the dispersion of the contributions from different companies. Instead of using $P(x_i)$ to represent the probability of the occurrence of event x_i , we use $P(n_commits_c^p)$ to represent the contribution ratio (declared above) of company c to the project p . Then we define the domination entropy of commits as follows.

$$CommitsEntropy^p = - \sum_{c=1}^n P(n_commits_c^p) \log_2 P(n_commits_c^p)$$

3 RESULTS

RQ1: Does One Company Dominate the Project's Development?

We found that the distribution of the contributions made by all the companies in each of the four projects is highly unbalanced. Thus, we form the following hypotheses to locate the dominant companies⁴ [6]:

¹<http://activity.openstack.org>

²[https://en.wikipedia.org/wiki/Gerrit_\(software\)](https://en.wikipedia.org/wiki/Gerrit_(software))

³<https://www.openstack.org/community/members>

⁴The dominant companies refer to a set of companies that make substantial contributions to the project.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

ICSE '18 Companion, May 27–June 3, 2018, Gothenburg, Sweden

© 2018 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-5663-3/18/05.

<https://doi.org/10.1145/3183440.3195047>

- The dominant company must contribute more than any other company.
- If one company makes more than 80% of the contributions of the company with the most contributions, we consider it to be a dominant company too.

Based on these hypotheses, we can obtain the number of companies that dominate the development of the four projects respectively.

Figure 1 shows a visualization of the contributions made by the dominant companies in the four projects per version respectively. The color depth of the rectangle represents the magnitude of the domination intensity (defined in Section 2.2) of the dominant companies. The number in the rectangle represents the number of dominant companies in the current version of one project. The “*” represents the change of the company that made the greatest contributions from the previous version.

We can see that, there are nearly half of the versions, in which the four projects are dominated by one company. The dominant companies change occasionally, but the number never exceeds three in all versions of the four projects. For example, the top dominant company switches between Rackspace, Red Hat, and IBM for Nova project.

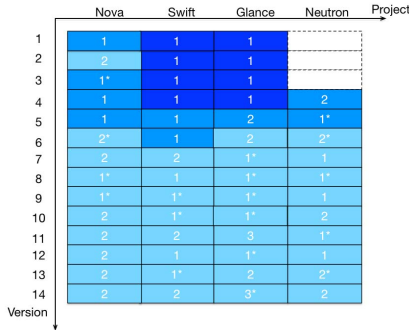


Figure 1: Visualization of the dominant companies of the four projects per version

RQ2: Does the Domination Affect the Community?

We conduct Spearman correlation tests to explore the association between two variables [10]. In our tests, one variable is a metric for company's domination, including the domination intensity and domination entropy for commits. The other variable is one of the metrics for the FLOSS community, including the number of companies and developers, the average productivity of the developers, and the quality of issue reports (the ratio of the number of fixed issues to the number of reported issues).

All of our tests have a p-values less than 0.05, indicating a statistically significant relationship [5]. More specifically, the domination intensity have a negative association with the participation of both developers and companies. This might be because the minor companies are concerned about being free labor to the dominant company and therefore are reluctant to participate [11, 13]. Meanwhile, the domination intensity has a positive correlation with the average productivity of the developers. This might be because the dominant company tends to set up a special team to speed up commit reviews and set reasonable milestones in the project. Furthermore, the positive correlation between the domination intensity and the quality of

issue reports suggests that the dominant companies might take special measures to promote the development. This is consistent with the comments from an expert in OpenStack: "In order to ensure the quality of the software, the dominant company generally provides a team for testing, so the issue reports reported by these developers are usually of high quality." In contrast, when the contributions are more dispersed (within a higher domination entropy), more companies and contributors are involved in. However, the domination entropy has a negative relationship with the average productivity of the developers and the quality of issue reports. These reasons above can also explain the correlation between the domination entropy and all the community metrics.

4 CONCLUSION

This paper conducted an empirical study on four main projects of OpenStack to explore the companies' domination in FLOSS. We found that one company often leads the project and the lead often switches between two companies. We found that the commercial domination is negatively associated with the participation of companies and contributors, while it is positively associated with the productivity of contributors and the quality of issue reports. Our study contributes to the understanding of commercial domination and its impact on FLOSS community. Our findings reveal the adverse consequences brought by commercial domination in addition to its benefits.

5 ACKNOWLEDGMENTS

This work is supported by the National Basic Research Program of China Grant 2015CB352200, the National Natural Science Foundation of China Grants 61432001, 61421091, 61620106007 and 61690200.

REFERENCES

- [1] Christian Bird, Nachiappan Nagappan, Brendan Murphy, Harald Gall, and Premkumar Devanbu. 2011. Don't touch my code! examining the effects of ownership on software quality. In *ACM Sigsoft Symposium and the European Conference on Foundations of Software Engineering*. 4–14.
- [2] Mathieu Goeminne and Tom Mens. 2013. A comparison of identity merge algorithms for software repositories. *Science of Computer Programming* 78, 8 (2013), 971–986.
- [3] Dietmar Harhoff, Joachim Henkel, and Eric Von Hippel. 2003. Profiting from voluntary information spillovers: how users benefit by freely revealing their innovations. *Research policy* 32, 10 (2003), 1753–1769.
- [4] Audris Mockus, Roy T Fielding, and James D Herbsleb. 2002. Two case studies of open source software development: Apache and Mozilla. *ACM Transactions on Software Engineering Methodology* 11, 3 (2002), 309–346.
- [5] Regina Nuzzo. 2014. Statistical errors. *Nature* 506, 7487 (2014), 150.
- [6] International Organization Standardization. 1995. Guide to the Expression of Uncertainty in Measurement. (1995).
- [7] Jose Teixeira, Salman Mian, and Ulla Hytti. 2016. Cooperation among competitors in the open-source arena: the case of openstack. (2016).
- [8] Joel West and Scott Gallagher. 2006. Challenges of open innovation: the paradox of firm investment in open-source software. *R&D Management* 36, 3 (2006), 319–331.
- [9] Wikipedia. 2017. Entropy (information theory). [https://en.wikipedia.org/wiki/Entropy_\(information_theory\)](https://en.wikipedia.org/wiki/Entropy_(information_theory)). (2017).
- [10] George Udny Yule. 1919. *An introduction to the theory of statistics*. C. Griffin, limited.
- [11] Yuxia Zhang, Minghui Zhou, Wei Zhang, Hanyan Zhao, and Zhi Jin. 2017. How Commercial Organizations Participate in OpenStack Open Source Projects. *Journal of Software* 28, 6 (2017), 1343–1356.
- [12] Minghui Zhou and Audris Mockus. 2012. What make long term contributors: Willingness and opportunity in OSS community. In *Software Engineering (ICSE), 2012 34th International Conference on*. IEEE, 518–528.
- [13] Minghui Zhou, Audris Mockus, Xiujuan Ma, Lu Zhang, and Hong Mei. 2016. Inflow and Retention in OSS Communities with Commercial Involvement: A Case Study of Three Hybrid Projects. *ACM Transactions on Software Engineering & Methodology* 25, 2 (2016), 13.