Addition of JSON Schema Keyword "allOf" in OPA Type Checker and the Overall Importance of Open-Source Projects
(Technical Paper)

An Exploration of the Benefits and Drawbacks of Open-Source Projects
(STS Paper)

A Thesis Prospectus Submitted to the

Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements of the Degree
Bachelor of Science, School of Engineering

Julia Friedman

Fall, 2022

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Introduction

It is widely known that open-source projects are very important and beneficial for developers and organizations alike. Open-source projects provide source code that is free and may be copied or altered by anyone (Corrado, 2009). All open-source projects can thus undergo perpetual enhancement by developers anywhere and at any time. Von Krogh and Spaeth (2007) present the advantages of the accessibility provided by open-source projects, including its value in promoting research. Essentially, it opens the door to an abundance of data for researchers and opportunities for developer improvement of code. Open-source also allows for continuous improvement of products and services, and for an open conversation about the functionality and enhancement of these software. A well-known example of an open-source product is the programming language, Python. Python is available for all users for free, allowing it to be easily distributed and broadly accessible. A user can download any source code within Python and further modify and distribute their version of that code. Overall, open-source encourages collaboration, community, and transparency among developers.

Although open-source software has many benefits to its users, there can be drawbacks of implementing such software. Open-source code can compromise security of applications, lacks the level of service and support that comes with commercial software, and introduces a risk of orphan software (*Pros & Cons of Open Source in business,* 2021). Regarding security, because open-source is not developed in a controlled environment and anyone can alter the code, there is a chance that a developer could have malicious intentions and compromise the security of an organization's data. It is also very possible for a well-intentioned developer to introduce a vulnerability into the code base by mistake. It is important for organizations that decide to utilize open-source software to be conscious of the cybersecurity risks involved. In terms of support, while commercial vendors

offer continuous technical user support for their software, open-source products do not provide such services. As a result, organizations may have trouble managing their use of open-source software, especially if they lack the necessary technical skills. This issue may cause a company to have to spend more money and resources on learning the necessary skills for software servicing. Finally, orphan software is software in which the developers have lost interest in the maintenance of the project. The software therefore becomes outdated or faulty due to a lack of upkeep (*Pros & Cons of Open Source in business,* 2021). The motivation for my overall technical research is to uncover the benefits and drawbacks of open-source in order to determine how companies may decide what kind of software best fits their needs.

Technical Topic

My interest in open-source projects was piqued during an internship where I completed a project that involved an enhancement of the type checker for one of Styra's open-source products, Open Policy Agent (OPA). OPA is an open-source engine that unifies policy enforcement across the cloud native stack, providing a declarative language that allows for specification of policy as code. However, one of the limitations of its Rego type checker was the absence of the keyword "allOf." My technical work involved adding support for the implementation of this keyword, which enhanced OPA's type checker, as users now receive detailed error messages regarding the use of this keyword. The three main components of the design of this project were comprehension of OPA's code base and its limitations, using JSON schemas for code structure validation, and understanding the relationship between OPA's Rego type checker and its error messages.

The first steps of this project were reading and determining what each part of the code does. The limitations of the code were already defined in the Issues tab within the OPA GitHub

repository, and one of these issues was the absence of implementation for several keywords in the type checker. The goal of this project then became adding support for the “allOf” keyword.

In order to implement a new keyword, it was necessary to understand how the other keywords are defined and tested and further to understand in what location in the code itself keyword definition and testing was implemented. Achieving this understanding relied on coupling the use of JSON schemas and OPA’s type checker. Schemas are a helpful feature of policy management systems in order to perform type checking. A JSON (JavaScript Object Notation) schema is a tool for validating the structure of data, outlining what input should look like. Schemas are typically passed in as input when evaluating a piece of policy code, and the associated type checker will use that schema as a blueprint to authenticate the structure of the code and inputs. In the case of OPA, the JSON schemas passed in as input during evaluation can interact with code written in Rego, which is a declarative language utilized by OPA for policy writing. When these schemas are included in the evaluation input, the Rego type checker uses them as a guideline for the intended code structure.

With a JSON schema as a guideline, the type checker can subsequently give more detailed and helpful error messages, as it can compare the usage in the actual policy code to that defined in the schema. The type checker gains the ability to give error messages that indicate if the user has made a mistake involving *those particular* implemented keywords. For example, if a developer made a typo involving the use of “allOf” in their inputted schema, the type checker could tell them that the root of that error relates to “allOf”, instead of generally telling them that there was an error. Developers leveraging OPA can still use keywords that are not yet implemented in the type checker, but any type errors that they make in using such keywords would result in difficult-to-debug output.

The final steps of this project were to combine the knowledge of the code base, JSON schemas, and OPA's type checker in order to create a solution and add code for implementation of "allOf." After testing, this code was submitted as a Pull Request into the main branch of the OPA GitHub repository; this Pull Request was approved by the maintainers of OPA and was thus incorporated into the open-source, publicly available code base.

STS Topic

The general topic area of this paper is the nature, value, and risks of open-source projects, and how this technically connects to the project I completed during my summer internship. Since OPA is an open-source project, it is readily available and free to those organizations who want to use it for their needs, whether it be for personal or commercial application policy management. In fact, many well-recognized companies use OPA, including Google, Netflix, Atlassian, and Capital One (Open-Policy-Agent, 2022). My project, which involves the continued enhancement of this service, will improve those companies' (and more) experiences using OPA. This topic is closely related to the topic of my technical thesis for computer science because this is the most significant code contribution I have made to date. I am also interested in the implications and benefits of the development of open-source projects.

The stakeholders of open-source software are people who care about the success of the product—these are traditionally the project's contributors and users. The main users of open-source projects are businesses that have a need to leverage technologies provided by these services. The stakeholders are not always involved the in the operations of open-source projects because they do not have an actual monetary share in the company (Peters, 2018).

The results of my technical project included a contribution into the OPA open-source project that enhanced its type checker by adding support for an additional keyword. The

completion of this project helps OPA as an organization because it is an instance of the continuous improvement that is possible because OPA is open-source. The creators of OPA did not have to do any work (aside from approving the Pull Request in GitHub) in order to improve their product.

This work is also important for the many organizations that utilize OPA for their policy management needs. Companies such as Google, Netflix, Atlassian, and Capital One all use OPA for their business needs; these companies benefit because the software that they are using has now been enhanced. The specific realized improvement allows for easier and more immediate type checking regarding the use of "allOf," and thus provides more precise and informative error messages for users in these organizations who write policy using this keyword. This enhancement saves developers time and effort in debugging, and saves their employers money.

Companies that utilize open-source projects like OPA must still be mindful of the associated risks. In the case of OPA, companies should consider the fact that because anyone can alter that software, it is not a 100% secure software. I was able to modify the source code, and I was merely a twenty-year-old summer intern. In order to alleviate this risk, open-source software managers and users can do thorough security checks with respect to the safety of the open-source code they are using.

Research on the tradeoffs of using open-source is important and worthy of attention because open-source can be a great decision for a company, but it is not always the right decision for every company's needs. Organizations need ample information in order to make business decisions, and this research can contribute to their understanding of both the positive and the negative implications of using open-source technologies.

The main question that I plan to answer through my research is: how should an organization evaluate the benefits and drawbacks of utilizing open-source products in order to fit their business

needs? I plan to answer this through extensive research on the tradeoffs involved in using open-source. This research will have a focus on what business circumstances may lead a company to want or need different aspects of the tradeoffs of open-source. I will also explore potential advantages and disadvantages of using OPA specifically, so that I can relate my STS discussion back to my technical topic.

Conclusion

Through my technical and STS deliverables, I hope to make meaningful contributions to both the open-source community as well as the potential users of open-source products. My technical project will benefit companies that utilize OPA for their policy management needs; they will have a more efficient debugging process, as I implemented an enhancement to the type checker that improves debugging error messages. My STS deliverable aims to benefit companies that are unsure of whether using an open-source product is the most tactical business decision for them. Such companies can evaluate the use of open-source through my analysis of the benefits and drawbacks of the nature of these products.

References

Corrado, E. M. (2009, June 16). *The importance of open access, open source, and open standards*. Retrieved September 26, 2022, from http://codabox.org/15/1/istl.pdf

Krogh, G. von, & Spaeth, S. (2007, August 13). *The open source software phenomenon: Characteristics that promote research*. The Journal of Strategic Information Systems. Retrieved September 26, 2022, from https://www.sciencedirect.com/science/article/abs/pii/S096386870700025X

Open-Policy-Agent. (2022, August 11). *OPA/adopters.md*. GitHub. Retrieved September 26, 2022, from https://github.com/open-policy-agent/opa/blob/main/ADOPTERS.md

Peters, S. (2018, July 18). *Do you know who your stakeholders are? - o'reilly open source convention in Portland 2018*. Open Source Conference | O'Reilly OSCON. Retrieved October 31, 2022, from https://conferences.oreilly.com/oscon/oscon-or-2018/public/schedule/detail/66945.html#:~:text=Stakeholders%20are%20the%20people%20that,a%20project's%20contributors%20and%20users.

Pros & Cons of Open Source in business. Investintech. (2021, February 26). Retrieved October 31, 2022, from https://www.investintech.com/resources/blog/archives/7975-pros-cons-open-source-business.html