Muhammad Salman Abid — Statement of Purpose for Ph.D. in Computer Science

1 Introduction

One of my favorite quotes in Computer Science literature "Andy giveth, and Bill taketh away" [2], also known as Andy and Bill's Law, encapsulates a challenge that I am particularly passionate about. Incremental gains in silicon performance are often nullified over time by newer software, and this challenge presents an ever-present opportunity for improvement in software systems. I regard this as a problem quite worth solving.

My research interests are broadly in Software Engineering and in Computer Systems. I am particularly passionate about building software tools, especially those that enable efficient and reliable compute. Through my research, I want to allow people to use software without being held back by reliability or performance concerns.

2 Research Background

My first experience with research was at Habib University, Pakistan, where I wrote my undergraduate thesis with Dr. S. Saleha Raza, in automating creative generation of textile patterns. Habib University's liberal arts education places special emphasis on research in relevant socio-economic areas, which subsequently informed my thesis topic. Textiles are important in Pakistan in a social context, and patterns and colors are especially significant in the identities of local culture. Our research was focused on enabling local artists to accelerate their design process by prototyping faster, and was accepted as a novel application of AI in Art and Design at EvoMUSART'23 [1] [2]. My experience while working on my thesis was the driving factor behind seeking out further opportunities in research.

I later joined a program for undergraduate research in Software Engineering with the University of Illinois at Urbana-Champaign, supervised by Prof. Darko Marinov (UIUC) and Prof. Saikat Dutta (then UIUC, now Cornell). I joined their program as a part of a "mini-crowd": constituting a group of 32 students, all ultimately collaborating towards a *single* research goal.

Working with this group allowed me to interact with a diverse group of students from over 6 countries, and collaborate with people from a variety of backgrounds and interests. My work with the program can be divided into two parts: (i) individual research on large-scale code translation and (ii) (mini) crowd research on reliable validation of code translation.

For the first few months, I largely studied the feasibility of code translation of a large microservices project (specifically Cornell's DeathStarBench) and translated part of the project from Go into two different PLs: Node.js and Rust. After my initial translation in Node.js, I proposed an alternate translation in Rust – a translation that would be more equitable in terms of performance when compared to the original. To validate my translation efforts, I was able to report performance results for translations in both Node.js (0.4x speedup) and Rust (1.14x speedup) when integrated into the original DeathStarBench cluster.

For the second part of the program, we worked on testing the validity of code translations in a more reliable manner; translated tests may produce misleading results due to differences in language-specific constructs. More specifically, we investigated a novel method of first (i) translating Java code for the Apache Commons CLI library into Python and then (ii) creating

[1] https://msalman-abid.github.io/publications/FabricSketchAugmentation.pdf

[2] https://github.com/msalman-abid/Project_Fabricate

bindings between Python and Java to allow tests from the *original* language to be executed on the Python translation. Employing this technique would leave no ambiguity in both the validity and reliability of the Python code.

The code for these bindings, or "glue code", was written using the GraalVM framework. I led this part of the translation, with more than 50% commits authored in the final version while collaborating with more than 10 other students on the project. Some of my work led to an open-source contribution in GraalPy, as I came across unexpected behavior in its runtime while interfacing Java and Python. I was also responsible for assisting the program lead(s) in preparing the necessary scripts and documentation to ensure reproducibility for our artifacts. The final manuscript for this work is under preparation.

In addition to my technical contributions, I also contributed to a teaching study of our unique structure of mini-crowd research. The structure of our program was inspired from the original Crowd Research Initiative at Stanford [1], which had over 1500 participants over a period of two years. In contrast, our program initially had 32 participants with an approximate duration of only 5-6 months.

The idea behind the program, and particularly its remote nature, was to allow more researchers from under-represented backgrounds to have access to research opportunities with professors who usually engage with undergraduates only at their own institutions. We prepared and submitted a manuscript describing our experience to the SEET track at ICSE'24, where it is currently under review.

Experiences such as ours also enable upward mobility in both education and career opportunities for students, for which I can personally vouch. I was able to contribute towards research while still working a full-time job, and also discovered my affinity towards research in software engineering in the meantime.

3 Conclusion

Having been involved in broader research for the past year, I am interested in exploring aspects of systems and software that need reliable and performant components, especially in the context of cloud computing. With my past experience in software testing, I am increasingly convinced that reliable testing is key to writing dependable software. This conclusion also translates well to lower-level aspects of computing, such as ensuring consistent and fault-tolerant behavior of operating systems.

As such, I am excited at the prospect of working with several professors at Cornell that explore these topics. With Prof. **Saikat Dutta**, I would want to continue working with him on building performant ports of large-scale ML libraries. My work with him has included feasibility studies for language bindings for Python ML libraries, which can be extended into large-scale translations that port these libraries to other languages, such as Java or JavaScript.

I would also like to work with Prof. **Owolabi Legunsen** in the intersection of Software Engineering and Systems, such as his work on cloud operator correctness with Acto. I could contribute to their effort on making Acto compatible with multiple interdependent operators, extending its capabilities beyond only supporting individual operators, to make it practical for real world applications.

Finally, I am also interested in working with Prof. **Adrian Sampson**'s group, Capra, on their current research with "Filament". I may be able to contribute to their efforts in expanding

their support for Filament, especially in generating hardware accelerator pipelines for data-dependent timing behavior.

For my doctoral studies, I also find Cornell's requirement for an external minor to be particularly beneficial. Given the chance, I would pursue a minor in Applied Statistics to allow me to complement any performance analyses I carry out to have a strong grounding in statistical significance, and construct a well-rounded approach to my research.

Beginning my academic journey with any of the faculty members of interest would be an experience that I would cherish, as I believe it would be a precursor to work that would have real-world impact. After obtaining my PhD, I hope to both continue working towards my research interests, and to mentor aspiring researchers (especially from under-represented communities).

References

[1] Crowd research initiative. `http://crowdresearch.stanford.edu/initiative`, 2017. Last accessed 03 November 2023.

[2] SUTTER, H. The free lunch is over: A fundamental turn toward concurrency in software. `http://www.gotw.ca/publications/concurrency-ddj.htm`, 2004. Last accessed 03 November 2023.