Statement of Purpose

Mina Lee
https://minalee.info Computer Science

My research objective is in the area of programming languages and computer security. Through my research, I would like to leverage program reasoning and synthesis techniques to identify and eliminate security vulnerabilities in programs. To this end, I hope to pursue a Ph.D. in Computer Science at Stanford University.

1 Research Experience

I worked as a research intern at the Programming Research Laboratory at Korea University. Participating in various projects on program synthesis, static analysis, and machine learning gave me the opportunity to become better acquainted with the essentials of programming languages.

Regular Expression Synthesis. In my joint graduation project, we built a system to help students learn how to construct regular expressions [1]. Given positive and negative examples, it automatically constructed the simplest regular expression consistent with the examples. To develop the system, I formulated the synthesis task as a search problem, designed an efficient search algorithm which exploits the semantic properties of regular expressions, incorporated pruning techniques, proved the soundness of the algorithm, and made it available online [2]. This experience provided me a good, overall understanding of program synthesis as well as theoretical and practical programming language techniques. This work earned us the *Best Paper Award* from the International Conference on Generative Programming: Concepts and Experiences and the Gold Prize in the Graduation Project Competition at Korea University.

Selectively Unsound Static Analysis. This work aimed to reduce the number of false alarms in static analysis, specifically interval analysis for buffer-overflow detection. Our insight was that the parts of programs that tend to produce only false alarms share common characteristics, whereas real buggy parts are unpredictable. We used a one-class support vector machine to learn such properties from the result of the analysis, and used them to selectively analyze new programs unsoundly. It was the first project I participated in, and my role was to reform the work based on the reviews from previous submissions. It was an instructive experience in that I learned the qualities required for a paper to be properly organized and supported, thereafter successfully completing ensuing projects.

Automatic Feature Construction. Another project was on generating features from a codebase for machine learning-guided adaptive static analysis. The motivation of this project stemmed from the burden on domain experts to create a set of features for each instance of static analysis. In brief, our approach performed static analysis on programs, automatically extracted and refined code snippets based on the result of the analysis, and selected a subset of them to serve as feature programs. In the project, I examined the output of the system and reasoned about their representativeness and effectiveness as features.

In addition to these projects, attending international conferences, PLDI and OOPSLA, was the culmination of my academic life. In presenting posters and sharing opinions, I could narrow down my research interests and solidify my ideas. At school, I audited graduate courses such as *Program Analysis* and *Programming Language Theory* as well as an undergraduate course *Programming Languages*. Also, I independently studied SAT/SMT solvers to familiarize myself with satisfiability-based technologies. These numerous undertakings further broadened my perspective and strengthened my background in programming languages.

2 Research Interests

My primary interests are program synthesis, reasoning, and repair. I would like to apply these techniques to interdisciplinary domains, spreading programming language ideas beyond the field.

Program synthesis has the enormous potential to make our lives more convenient by automating complicated, demanding, and repetitive tasks. To improve the current synthesis algorithms, I will leverage a range of techniques including sketching, constraint solving, and machine learning. With learning-based approaches, I would like to mine information from existing repositories such as GitHub to facilitate the synthesis process. I will also strive to develop more effective and scalable program synthesis techniques through modular and incremental synthesis. At the same time, I hope to bring these theoretical insights to pragmatic use by building tools and applying them to a variety of domains such as education and security.

Education. I aspire to build a personalized educational assistant that tailors itself to the needs of each student at both the university level and before, providing greater learning opportunities to reach more people. For programming assignments, this could involve designing domain-specific languages for correction strategies, reasoning about symbolic equivalence of programs, inferring program distances for quantitative repair, and automating error localization and correction of programs.

Security. Automated reasoning and program repair can find and remove security vulnerabilities in programs. In particular, I would like to explore the use of SMT solvers for software security such as vulnerability checking and exploit generation. Furthermore, I hope to use program synthesis techniques to generate a set of syntactic changes which can remove the vulnerabilities. I believe my former involvement in computer security from taking the class *Introduction to Cryptology* to participating in extracurricular activities in *Korea University Institute of Computer Security* would serve as starting points for applying programming language-based techniques in the context of security and privacy.

3 Conclusion

With this focus and potential directions to keep in mind, Stanford University would be an ideal place to pursue my doctoral study. In particular, I hope to work with **Professor Clark W. Barrett** as my research interests and area of expertise fit well in line with his research interests, specifically automated reasoning and satisfiability-based techniques. His expertise in the development and application of solvers for SMT will provide a rich context for my proposed research. Also, I am eager to receive the tutelage of **Professor John C. Mitchell** whose research problems are at the intersection of programming languages and computer security. His paper on massive open online courses [3], which highlighted opportunities to utilize their data for research purposes as well as concerns on the users' privacy, made me more conscious about the necessity of protection measures to lower potential security risks.

Upon completing my Ph.D., I would like to remain in academia as a professor to continue researching and teaching students. I firmly believe my proposed research can contribute to your university and academic community in that it will bring the findings and lessons learned from the research to campus courses as well as state-of-the-art technical support to millions of people concerned with their security. For these reasons, I am confident that Stanford University is a great place for me to pursue a Ph.D.

References

[1] **Mina Lee**, Sunbeom So, and Hakjoo Oh. 2016. Synthesizing regular expressions from examples for introductory automata assignments. In *Proceedings of the 2016 ACM SIGPLAN International Conference on Generative Programming: Concepts and Experiences* (GPCE 2016).

[2] AlphaRegex: Automatic Synthesizer for Regular Expressions from Examples. `http://prl.korea.ac.kr/AlphaRegex`

[3] Pierre Dillenbourg, Armando Fox, Claude Kirchner, John C. Mitchell, and Martin Wirsing. 2014. Massive open online courses: current state and perspectives. *Dagstuhl Manifestos* 4(1).

The format of this document is adapted from Anant Bhardwaj's Statement of Purpose.