Research Statement

Farzin Houshmand

Summary

I conduct research on applications of formal methods in programming languages that enable programmers to build safer, more efficient, and more reliable computing systems – in particular, subtle distributed and parallel systems. Before joining the Ph.D. program at University of California, Riverside, I finished my bachelor's degree in Computer Engineering at Sharif University of Technology. During my Ph.D., I have published number of papers in different venues such as POPL, ICFP, CAV and S&P. Meanwhile, I obtained my master's degree in 2021 in Computer Science from the University of California, Riverside.

Background

In the ever-evolving technological landscape, it is important to ensure that our increasingly complicated systems—including phones and computers—are reliable, safe, and fast. Formal methods, which are mathematically rigorous techniques for the specification, development, and verification of software and hardware systems, are a key consideration in the field of computer science because of this very concern. Performing mathematical and logical analyses can help make the design of a system more reliable and safeguards businesses from unwanted incidents such as data breach or malicious attackers. Here, I summarize my projects:

Distributed Systems Synthesis

Distributed systems and replicated data stores are integral part of today's modern computing platforms. Many companies replicate data on multiple servers across the globe to improve reliability, fault-tolerance, or accessibility of the service. To maintain the consistency of data, replicas need to participate in coordination protocols. To this end, developers get to choose from a spectrum of consistency choices. Coordination protocols with strong consistency guarantee the same total order of transactions and provide integrity and convergence; however, since such protocols need consensus between replicas, they may not be responsive or available during network failures or offline use. On the other hand, coordination protocols with weak consistency notions benefit from responsiveness and availability, but without guaranteeing the same total order of transactions. Despite the advancements in distributed systems technologies, developers still have a hard time understanding the numerous and low-level consistency criteria, and ultimately have difficulty choosing the optimal coordination protocol for their use.

In this work[1], I sought to understand the complications and challenges of choosing the right consistency model and ultimately develop a general framework to facilitate and automate consistency choices. Using a novel sufficient condition called well-coordination, I defined notions of conflicting and dependent pairs of methods and developed a static analysis technique that automatically analyzes the given object to infer the minimum required coordination that guarantees user defined integrity. By offering a synthesis tool that automatically finds the right consistency model as well as replication protocols that guarantee integrity and convergence, Hamsaz, had fundamental contributions to the correctness of commonly used applications. Later, in a follow-up work[2], we extended Hamsaz with another favorable property in replicated systems, namely, data

recency which is often neglected in the literature. Hampa provides additional interface for the users to specify the recency requirements for the methods of the object and automatically offers a replication protocol that not only preserves the integrity and convergence, but also respects the user defined recency requirements.

The systems above assume a failure model in which nodes of a distributed system can fail only by crashing. However, building trustworthy systems in the face of malicious attacks demands significant consideration. Inter-organizational systems where subsystems with partial trust need to cooperate are common in various bussinesses. In the face of malicious Byzantine attacks, the ultimate goal is to assure three aspects of confidentiality, integrity and availability. In contrast to confidentiality and integrity, studying availability policies has been often neglected. In my next work, I regard this shortcoming. This paper[3], presents a security typed object-based language along with an information flow type inference system that guarantees confidentiality, integrity and availability policies. Given a class and the specification of its policies, the Hamraz tool applies type inference to automatically place and replicate the fields and methods of the class on Byzantine quorum systems, and synthesize trustworthy-by-construction distributed systems.

Language Design

Companies and organizations around the world are increasingly using graph analytics to help uncover insights in numerous fields, including marketing, fraud detection, supply chain, search engine optimization, and more. Therefore, graph analytics must be able to run on graphs with millions of edges and vertices and as a result, different high-performance graph processing frameworks have been proposed. With that said, current graph analytics methods could use improvement, as they are often complex, platform dependent, and with manual optimizations being prone to error as well as time-consuming. In this project[4], I leverage formal methods and language design to simplify the error-prone and time-consuming task of graph analytics. This work offers a novel declarative language for graph analytics and include a set of transformation rules that optimize the specification. Given the specification of analytical problem, the tool can synthesize and generate optimized low-level code for different high-performance frameworks which can be used in many areas of computer science such as machine learning and data mining to help the programmer to write concise, efficient, and—more importantly—correct graph analytics. The resultant tool has been made publicly available[5].

Blockchain And Cryptocurrency

Recently, blockchain technology and cryptocurrencies have attracted attention of big tech companies and financial institutions and as a result, they have been the target of many attacks. Previous incidents on these systems show that the current infrastructure could use improvements in their safety and security measures. One application of rigorous reasoning is to reason about atomicity of the transactions that happen across blockchain. Although a transfer within a blockchain is atomic, the individual transfers of an exchange across blockchains are not atomic. Atomic execution of an exchange is challenging since single transfers are immutable and irreversible. Industries have been using protocols as-is without considering their correctness and safety guarantees. Cross-chain transactions can be represented as a directed graph with vertices as parties and edges as asset transfers. In a simple form, cross-chain transactions are cross-chain swaps where each edge e transfers an asset form source to the destination. Previous works introduced a uniform protocol for strongly connected cross-chain swaps and showed that no uniform protocol exists for transactions that are not strongly connected. However, in general, a cross-chain transaction includes a sequence

of exchanges at each blockchain. Further, transactions may have off-chain steps and hence may not be strongly connected.

This project proposes a new protocol for general cross-chain transactions with sequenced and off-chain steps. When a few certain parties are conforming the protocol guarantees the following properties: if all parties conform to the protocol, all the assets should be transferred. If any party deviates from the protocol, the conforming parties should not experience any loss. And finally, if the source parties pay, the sink parties are paid. We present a synthesis tool called XCHAIN[6] that given a high-level description of a cross transaction can automatically generate smart contracts in Solidity for all the parties.

Research Agenda

The type of research that I do makes the programming of distributed and concurrent applications easier and safer. I am planning to do more fundamental research in this area and make such systems even safer, faster, and more secure to prevent financial loss. Here I briefly touch on my short-term and long-term research plans. First, I describe my short-term agendas:

In the course of my Ph.D. I realized there is still a considerable gap between state-of-the-art technologies in database systems and the opportunities for automation in this area. In the near future, my goal is to design and implement a fully automated database on top of the theoretical foundation laid by[1]. The first step is to design a language to capture common database programs. My plan is to formally define a language for relational algebra (the mathematical theory for modeling data) and characterize the decidability of the analysis method for this new language. This is an important work because it concisely captures which programs can be automatically analyzed and benefit from this approach. For that, I will work on a new decision procedure for relational algebra. Previous works have been studied decision procedures for the set theory, the theory with commonalities with relational theory. However, decision procedure for relational algebra has never been studied before. The decision procedure for different domains are at the core of SMT solvers which make automation possible. Proposing new decision procedure requires extensive study about the mentioned theory. Also, implementing the new procedure in existing SMT solvers require knowledge about solvers internal design. The result will be helpful for a wide range of audience in database community as well as researchers in programming languages.

For my long-term research plan, I want to apply my theoretical knowledge to make blockchains more secure and cryptocurrency transactions faster. Recent software bugs and hacks in blockchain networks have shown that these systems suffer from security flaws. Moreover, blockchain systems such as Bitcoin and Ethereum are slow by design. At the heart of any blockchain, there exists a consensus protocol which governs how participants agree on the state of the blockchain. As an example, Bitcoin is known to be wasteful when it comes to energy usage because of its proof-of-work consensus protocol. I plan to leverage my background in distributed systems and protocol design to come up with faster, more efficient, and more secure blockchain protocol.

Overall, my research will involve a good mix of theoretical and practical research. One part of my work will focus on fundamental foundations and radical solutions regarding reliability, performance, and safety of the computing systems. On the other hand, I intend to devote the other part of my work on building practical systems, which have immediate relevance and impact in Industry. I intend to work closely with researchers in related fields as well as Industry to understand, develop, and promote solutions for practical problems. I believe my experience of research work done jointly with my colleagues as well as my prior record of publications will help me achieve this. I am excited about the outlook of learning, contributing, giving shape, and making an impact in this challenging field.

References

[1] F. Houshmand and M. Lesani, "Hamsaz: replication coordination analysis and synthesis," *Proceedings of the ACM on Programming Languages*, vol. 3, no. POPL, pp. 1–32, 2019.

[2] X. Li, F. Houshmand, and M. Lesani, "Hampa: Solver-aided recency-aware replication," in *International Conference on Computer Aided Verification*. Springer, 2020, pp. 324–349.

[3] X. Li, F. Houshmand, and M. Lesani, "Hamraz: Resilient partitioning and replication," in *2021 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2021.

[4] F. Houshmand, M. Lesani, and K. Vora, "Grafs: declarative graph analytics," *Proceedings of the ACM on Programming Languages*, vol. 5, no. ICFP, pp. 1–32, 2021.

[5] "Grafs," https://zenodo.org/record/4968451, accessed: 2021-09-15.

[6] N. Shadab, F. Houshmand, and M. Lesani, "Cross-chain transactions," in *2020 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*. IEEE, 2020, pp. 1–9.