Statement of Purpose

Md Atiqur Rahman
Fall 2024: Computer Science, Ph.D.

In 2020, when the COVID-19 outbreak was at its peak, I along with some of my university peers came up with the idea of a relief activities coordination platform to assist numerous voluntary organizations in providing essential aid nationwide. Despite our dedicated efforts to launch the platform named Traan-Chitro, we encountered minimal interest from these organizations due to their concerns about data security and hesitance to adopt a new system. I was quite disheartened after this setback, but I never stopped thinking about why we failed. After some time, it became evident to me that our failure was not due to a lack of project potential, but rather stemmed from the oversight of conducting adequate research before initiating the idea. We overlooked assessing the targeted organizations' needs, their receptiveness to new systems, and their willingness to share data. That was the time when I learnt the critical importance of research, not just in understanding the technical aspects of software engineering and privacy, but also in acknowledging the pivotal role of human factors.

During my time at BUET, I developed a keen interest in software development and security. Engaging in challenging academic projects like creating 'E-luxurious'—a property rental platform like Airbnb—and designing a full-scale DNS flood attack with a corresponding defense mechanism heightened my skills. I also ventured into multiple outsourced projects to broaden my experience in software development practices. Later, I started participating in CTF competitions, where I learnt about various security tools. I especially loved solving reverse engineering problems and I found it amazing how much information can be recovered through this process. I came to know about this tool Ghidra, which is broadly used for software reverse engineering. I was fascinated by its decompilation capabilities. After graduation, I became a software engineer at IQVIA, specializing in C# .NET backend development. In this role, I extensively worked with various DBMS and gained proficiency in writing and analyzing unit tests and BDD tests in Gherkin language. I also focused on improving program analysis techniques to identify and resolve security vulnerabilities in developer-written code, ensuring compliance. These experiences allowed me to closely observe prevailing issues in the software industry, thus motivating me more to pursue a research career in which I can continue to delve further into software security. My ambition is to excel as a researcher specializing in **Security**, **Software Engineering**, and **AI**.

My journey into formal research began with my undergraduate thesis in computational criminology, where I developed a decision-aid system named 'Cri-Astrologer'. Its main purpose was to assist in the conduct of police investigations by predicting criminal demographic profiles using crime evidence data and victim demographics. With the guidance of my supervisor Dr. A. B. M. Alim Al Islam, I proposed a deep factorization machine based DNN architecture which outperformed existing machine learning and deep learning algorithms in predicting criminal demographics. It was published as a conference paper in **ACM NsysS'22**. At IQVIA, I have led research and development efforts to enhance database query performance, reduce query counts, and explore cost-efficient solutions. Currently, I am actively engaged in an R&D project at IQVIA aimed at revolutionizing user interactions with data visualization. The aim of this project is to simplify complex dashboard configurations by integrating large language models (LLMs) into the user interface. This approach will allow users to ask natural language queries and the LLM provides insights to automatically generated charts, delivering an exceptional user

experience. Throughout this project, I have explored various LLMs including GPT, Llama and Mistral, delving into prompt engineering and fine tuning these models. Working with LLMs in this project opened a new door of research interest for me: **Leveraging LLMs in solving software and security related problems**.

In pursuing my research interests encompassing **Security**, **Software Engineering**, and the **Application of LLMs**, I want to combine the strengths of each field to enhance one another. In my doctoral research, I aim to conduct an extensive investigation into software stack vulnerabilities, particularly examining the multifaceted applications of LLMs within the domain of software security research. This includes investigating and identifying ways to utilize LLMs in detecting and mitigating software stack vulnerabilities. Additionally, I am keen on enhancing existing security tools, delving into different static and dynamic analysis methods to develop practical solutions that can benefit the wider community.

I consider the Penn State College of Information Sciences and Technology a suitable place to pursue my PhD, as there are several active researchers with whom I believe I will be able to contribute. Regarding that, I am particularly interested in the **BRUCE** project at PIKE lab directed by **Dr. Dongwon Lee**. This initiative focuses on developing an advanced vertical search engine for cybersecurity learning resources, utilizing fine-grained Named Entity Recognition (NER) methods to extract detailed metadata. This aligns with my experience in leveraging LLMs for similar purposes in my ongoing research project at IQVIA. In this role, I've explored various LLMs to extract information from user queries, utilized for filtering different charts. Apart from LLMs, I've also worked with Amazon Lex and the spaCy library for similar tasks. I am confident that my expertise in software development and LLMs can contribute significantly to this project. I am also interested in collaborating with **Dr. Peng Liu**, given his focus on AI applications in cybersecurity. His recent work "**The Effectiveness of Large Language Models (ChatGPT and CodeBERT) for Security-Oriented Code Analysis**" piqued my attention. If given the chance, I am keen to explore how LLMs perform in identifying and analyzing vulnerabilities within software stacks using methodologies like '**fuzzing**'. I believe using LLMs to generate fuzzing inputs could potentially resolve several limitations inherent in existing fuzzers. Moreover, I want to explore different methods of detecting and preventing DNS cache poisoning attacks. Furthermore, I am also interested in collaborating with **Dr. Taegyu Kim**. His work on detecting phishing attacks in this paper '**Learning from the Ones that Got Away: Detecting New Forms of Phishing Attacks**' utilizes a machine learning classifier to detect phishing emails. As a future direction, I want to investigate how a fine-tuned LLM could perform in this context, leveraging its broad language comprehension. I am confident it will perform better than the benchmark tools. Additionally, I am also open to collaborating with others who share similar focus and exploring further areas in this domain.

My future goal is to become an accomplished academic, emphasizing both research and teaching while maintaining active connections with the industry. To achieve these goals, I am willing to explore new domains and embrace challenges that arise during my graduate studies. Please feel free to visit my portfolio at https://atiqur-rahman-0041.github.io/ for a detailed overview of my research, publications, and work experiences.