Research Statement

of Shoumik Saha (CSE PhD applicant for Fall—2022)

In this modern era of the digital revolution, electronic devices are no longer distinguishable from technology; rather they have become utilities, like water or electricity. As more and more people and devices are getting involved, everyday we are getting immersed with sensitive data generated from numerous sources, including social media, health monitoring devices, sensor networks, geospatial sources, genomics, financial, business informatics, and so on. Moreover, many advanced techniques from sophisticated domains (e.g., machine learning, operating system, network) are constantly being placed into these devices, making them more powerful. With great power comes great responsibility. So, this increasing number of devices, as well as users, deserves more security than ever. One attack or vulnerability on these devices can incur huge losses varying from a loss of billion dollars to the death of people. So, in my opinion, the most stimulating research topic is what strengthens this security. Thus, a Ph.D. in the field of Security, i.e., Computer Security, System Security, Network Security, Security with Machine Learning, is the logical outcome of my aspiration.

From my early childhood, I have always been fascinated and passionate about mathematics, logic, and I found computer science as a great opportunity to apply my mathematical skills and rational thinking. Therefore, I did my major in Computer Science and Engineering (CSE) from Bangladesh University of Engineering and Technology (BUET), which has an acceptance rate of 12% in the most competitive university of Bangladesh. A tremendously competitive environment during my undergraduate study motivated me to be more passionate and determined. The strong desire to establish my career in academia and research pushed me to secure a position among the top 19% of the students in my class (with 92% CGPA).

I got the opportunity to pursue my dream when I started working with Dr. Atif Hasan Rahman and Dr. Sadia Afroz on my undergraduate thesis focusing on malware security. Our aim was to classify malware families using a less resource-intensive model, with the best accuracy possible without compromising the robustness. We incorporated ideas and tools from bioinformatics with malware security. We used SibeliaZ, a multiple whole-genome sequence alignment tool, which is usually used to align DNA, RNA sequences. But in our novel approach, we modified this to align binary files. The capability of these tools to handle a significant amount of insertion, deletion and modification acted as the main incentive, and this feature helped us to make our model more robust and interpretable. Additionally, we used concepts of consensus sequence, conservation score and incorporated machine learning, back-tracking to construct the model. Our model was evaluated on the Microsoft Kaggle dataset and the Security Evision Competition dataset, and we compared it with three state-of-the-art models. In terms of accuracy, our model outperforms all these models, even the feature-fusion model (the paper with the best accuracy on Kaggle dataset to our knowledge). We proved the robustness of our model by demonstrating that no adversarial sample could evade ours, whereas some of these samples could evade one of the state-of-the-art models, MalConv. On top of that, due to its better interpretability, we could show some case studies where our method could locate the code blocks that were obfuscated by the attacker, and this feature can provide good insight to the security analysts. Another key strength of our method, Malign, is that it is scalable, less resource-intensive (does not require any disassembler) along with a competitive speed. Our research titled, 'MALIGN: Adversarially Robust Malware Family Detection using Sequence Alignment' has proceeded to the rebuttal period of IEEE Security & Privacy 2022 (acceptance rate: 14.9%). This research not only grew my interest in security but also taught me to look at a problem from a different perspective and be more open-minded.

In addition to security, my interest also lies in machine learning. As a research assistant in the Data Science and Engineering Research Lab, BUET, I am working on a Government-funded project detecting atrial fibrillation from noisy ppg signals in collaboration with Dr. Mohammed E. Ali, Dr. Mohammad M. Masud and Dr. Atif Hasan Rahman. Our initial approach was to apply Bayesian deep learning, and it outperformed traditional deep-learning models, including ResNext50 and other com-

plex models with more than 40M parameters. We developed an android app that takes ppg signals from smartwatches and alerts the user early in case of atrial fibrillation. Later we have extended this project using contrastive learning in an unsupervised way and incorporated poincare-plot. Thus, we reduced the false positive and false negative rates even more. Now I am working on preparing the manuscript and submitting the work to IMWUT for publication.

Besides my academic and research work, I have always kept myself involved in extra-curricular activities which helped me to enhance my communication, management and leadership skills. During my undergraduate studies, I worked as the general secretary of the photographic club at my university and organized exhibitions, workshops, seminars, etc. After my graduation, I joined as a lecturer at a renowned university, United International University, and I enjoy interacting and sharing ideas with my students there. Moreover, in our research lab, I am working on the above-mentioned project as a team leader where I organize regular meetings and monitor the progress.

In my junior year, I have taken various courses including Operating System, Computer Networks, Compiler, Computer Architecture, etc. to strengthen my base. During senior year, in the first class of Computer Security Sessional, we were shown the buffer overflow attack which caught my attention, and thus I got indulged in this course. I worked on different attacks and vulnerabilities in this course, such as Cross-Site scripting attack, TCP reset attack, DoS attack. Later I started working on my thesis on malware security, where I incorporated knowledge from my Bioinformatics and Machine Learning courses too. All these research and academic experiences have boosted up my confidence in pursuing a Ph.D. in my field of interest.

My vision is to actively contribute to the research and development of new tools and techniques in the security field, along with machine learning. I intend to work on strengthening the security of the technology sector of Bangladesh. Though my country is trying its best to cope up with technological advancement, it is still facing barriers to secure its services from attacks as a third-world developing country. For example, in 2016 illegal transfer of US$1 billion was issued by security hackers from the central bank of Bangladesh (Bangladesh Bank cyber heist), and this trend is still continuing with other banks frequently. Such losses have an immeasurable impact on the economy of a developing country like ours. As a computer science graduate, I feel the necessity of meticulous research in computer security and the urge of helping my country to move forward. In a broader sense, my aim will be to produce quality research output targeted at the enhancement of human life.

Please feel free to visit my online portfolio at https://shoumiksaha.github.io/ to view my research, publications and other scholastic achievements.