Enabling Privacy-Preserving Biomedical Data Analytics in the Cloud and Across Institutions

Haixu Tang, Indiana University, Bloomington; *Xiaofeng Wang, Indiana University, Bloomington; Yan Huang, Indiana University, Bloomington; Xiaoqian Jiang, University of California, San Diego; Shuang Wang, University of California, San Diego; Lucila Ohno-Machado, University of California, San Diego*

The outsourcing of biomedical data into public cloud computing settings raises concerns over privacy and security. With respect to public cloud environments, there are concerns about the inadvertent exposure of human genomic data to unauthorized users. In analyses involving multiple institutions, there is additional concern about data being used beyond agreed research scope and being processed in untrusted computational environments. Significant advancements in secure computation methods have emerged over the past several years. A key challenge here, however, is how the analytical algorithms can be conducted in an efficient and cost-effective manner. Some of these techniques were implemented in general-purpose packages, they are not optimized to the analysis of biomedical data (note that these cryptographic algorithms are often computationally intensive, and thus the optimization to specific computational tasks can lead to acceleration of several magnitudes), and their applications are not straightforward. In our project, we attempt to develop a suite of methods and open source software tools that can be used by biomedical researchers in a plug-and-play manner for the statistical analysis of encrypted biomedical data. Our methods assume biomedical data will be protected by encryption after they are generated, and the subsequent analysis and sharing will always be performed on the encrypted form. We will utilize existing general-purpose encryption software, and will develop and optimize them for biomedical computation tasks. We will also evaluate these methods rigorously for their ability to support the analysis of various type of biomedical data. In the past two years, we organized twice the Critical Assessment of Data Privacy and Protection competition to assess the capacity of cryptographic technologies for secure human genomic computation in the cloud and cross-institutional collaborations. We will report our findings from the competitions and our future plan.