For office use only	Team Control Number	For office use only
T1 _____	93036	F1 _____
T2 _____		F2 _____
T3 _____	Problem Chosen	F3 _____
T4 _____	F	F4 _____

2018
MCM/ICM
Summary Sheet

PIPE: Estimate the Value of Private Information

Summary

Contrary to the pervasive belief that human society has entered the information age, the massive data produced by human individuals are not fully exploited yet. Private data nowadays are under poor, isolated management by individual enterprises, where the value of data cannot be fully extracted to benefit either its provider or owner. To address this problem, a well-established market system is required that not only prices and rewards data sharing, but also regulates and protects private information.

To satisfy the requirement, our paper provides a detailed analysis based on a dataset *PI-DATA*, based on which we propose a sophisticated and generalized model, Private Information Price Estimation (PIPE), which is able to estimate the price of private information (PI) regarding different data domains of PI and social subgroups.

Task 1: We abstractly extract *feature vectors* from individuals and query requests to distinctly characterize their traits in different data categories..

Task 2: We estimate the correlation matrix of data categories and develop an amendment formula to accurately compute data value considering internal and external factors.

Task 3: We establish a *Supply and Demand Model* to estimate the value of PI as a commodity on the level of individuals, groups and nations.

Task 4: We surveyed the existing government act (e.g. Privacy Act, GDPR, APPI, etc.) and price regulations related to the private information around the world. Also, we introduce a dynamic variation to illustrate the change of human decision-making over time.

Task 5: We introduce a risk-to-benefit factor and show how generational differences change our model. We also compare PI with PP and IP.

Task 6: To clarify the connection between different subgroups of people, the multi-dimensional clustering algorithm for friends (*mCAF*) is applied to the dataset *PIDATA*. By conducting experiments on the data from different groups as well as from the same group, we find that the relationship between data and value is not linear, but log-likelihood.

Task 7: We simulate the effect of massive data breach and predict the effect of PI loss and cascade event using our model. Based on our pricing system, we think agencies should compensate to individuals directly for data breaches.

In the end, we make sensitivity analysis and discuss the strengths as well as weaknesses of our model. Moreover, a policy memo is presented to the decision maker on the utility, results and recommendations based on our *PIPE* policy model.

Keywords: Private Information; Pricing Strategy; Dynamic System; Network Effect