PEDE: A Cloud-Based Personal Execution and Data Hosting Environment

Rayman Preet Singh, S. Keshav and Tim Brecht
University of Waterloo

Increasing amounts of data are being generated and collected by, on behalf of, and about individuals. Some of this data is generated by traditional applications and services like document processors, e-mail, media-sharing services, web browsers, instant messaging, and social networking services. Other emerging sources of data include devices that act as sensors to record data such as smart metering, heath-care monitoring, smartphone-based sensing, and monitoring of individuals' banking and shopping activities. Most often, data is collected by service providers who take ownership and full control of the data - thereby risking users privacy - in exchange for free services. There is a growing discomfort among consumers about relying on the service providers' changing privacy policies, losing data privacy and control, and having to trust these services. This is evident from dissent against leading social networking and media sharing services, and cases of serious user resistance to the installation of smart meters collecting energy consumption data. These concerns are not without warrant as recent research has demonstrated that such data can be mined to reveal private information about users. For instance, energy consumption data from smart meters can be used to determine occupancy, appliance use, and even the TV channel being watched! Other forms of user data such as messaging, photos, videos, location, health statistics, spending activities are unarguably private in nature and their collection by service providers poses new threats to user privacy. However, keeping user data completely private makes it impossible to make data driven recommendations to users that could benefit them. Our goal is to build an environment that balances data privacy and data analytics.

We propose constructing a framework in which users place their data at a universally accessible location that they own and control individually. A user's data resides in the cloud within her *Personal Execution and Data Environment (PEDE)* which provides reliable storage for hosting the data, and computation to run applications on the data. The use of modern clouds for hosting PEDEs relieves the user of the problems of warehousing the data, its accessibility, computation resources for its processing, and its consolidation from multiple sources, which arise when commodity devices are used for the purpose. Users download applications to their PEDEs and they interface with users' data and other services a PEDE may offer. Being a PEDE owner, a user can configure applications' access to the data (and services) and can impose her own privacy policies, enabling a privacy-preserving application ecosystem for the data, which remains under her purview at all times. In this ecosystem, third party developers build applications that process the data, generating meaningful results for the user, enhancing data value, while fully respecting users' data ownership, privacy and control. Much like *app stores* for mobile devices that have enriched the user experience, such an ecosystem would enable innovation in data processing which is currently frozen because of user data being locked with service providers.

We are studying a cloud-based architecture that uses PEDEs to allow users to provide third parties with fast, consolidated, universal, and privacy-preserving access to their data while retaining complete ownership and control of the data. We build example applications to demonstrate how appliance vendors, energy auditors, and other third parties can develop consumer applications using our platform while preserving consumer privacy. Possible use cases of this platform include: applications performing detailed analysis, tailored to individual users, for quantifying benefits of purchasing energy efficient appliances, and for helping users better understand and control their energy consumption.

Prior work has recognized the problem of data privacy and offered theoretical advances, such as differential privacy and homomorphic encryption. It provides protocols that protect the privacy of the data while enabling computations on that data. Unfortunately, prior work does not describe systems to enable application development and deployment. Our work is unique in that it leverages the rich infrastructure of modern clouds to provide an environment for the implementation of these algorithms.