Position Statement:
Towards a Secure and Flexible Personal Data Platform on the Edge

Tamer Nadeem
Virginia Commonwealth University, USA
tnadeem@vcu.edu

GEFI Participant: Tamer Nadeem

Research Description:

Internet of Things (IoT) is becoming the key enabler for highly intelligent data rich applications and is the major technology behind smart-computing domains like smart homes, connected health, connected cars, automated enterprise workflows, Smart Cities and Smart grid. Ericsson predicts the number of connected IoT devices to be around 18 billion by 2022 and will take a lead over the number of mobile phones in 2018. Increased ubiquity of sensing via smart devices and IoT devices in smart homes and smart healthcare domains, for example, has caused a surge in sensitive and personal data generation and use from browsing habits to purchasing patterns to real-time location to personal health information. Unfortunately, our ability to collect and process data has overwhelmed our ability to protect that information in which concerns over privacy, trust, and security are becoming increasingly important as different stakeholders attempt to take advantage of such rich data resources.

In our view, the core technical problem in this space is how to provide technical means that enable and assist individuals in managing access to their data by others while retaining personal control over such uses and understanding the implications of any data release. Moreover, the significant growth and penetration of smart and IoT devices come along with a tremendous increase in the number of smart and IoT applications. These various applications, which support various domains and services, generate and access different data patterns such as periodic, event-based, realtime and continuous data. These different applications result in diverse traffic characteristics that require different performance levels of reliability, loss, and latency. To cope with this various traffic characteristics and requirements, it is now necessary to have greater visibility and control over the traffic generated from smart and IoT devices in order to guarantee an optimized performance of smart and IoT applications as well as high quality of experience to users.

In this research, we aim to design and develop ExtremeDataHub platform an open-source, flexible, and programmable networked edge device that collates and mediates access to our sensitive and personal data, under the data subjects control as well as to cope with various characteristics and requirements of smart and IoT applications that access this data in order to provide better performance and quality of experience to users. ExtremeDataHub will leverage and combine the current project of our group SMILE - SMart and Intelligent wireLess Edge[1], which is framework that supports Software Defined Network (SDN)-like paradigm at user smart devices and network wireless-edges enabling network wireless-edges to become more active and providing truly end-to-end management and control in which users could reap the full benefits of SDN, and the current project of Dr. Hamed Haddadi (from Imperial College London UK) Databox[2], which is an open-source personal networked device augmented by cloud-hosted services enabling individuals to manage their data and to provide other parties with controlled access to their data, into a holistic integrated platform; ExtremeDataHub.

---

[1] https://music.lab.vcu.edu/smile/

[2] https://www.databoxproject.uk/

ExtremeDataHub will be developed as a networked edge device (e.g., could reside within the home wireless access point or as a standalone device) that enable individuals to control and manage their sensitive and personal data as well as to provide flexibility and controllability to support the various characteristics and requirements of different data access flows generated by smart and IoT applications for higher performance and quality of experience. Figure 1 shows the proposed ExtremeDataHub platform that runs on edge devices (e.g., WiFi access point, stand alone edge unit). ExtremeDataHub modules are divided to two main components: ExtremeDataHub data plane and ExtremeDataHub control plane. ExtremeDataHub data plane consists of both SMILE data plane (i.e., forwarding network components) and the three main data access components of Databox (i.e., driver, store, app). On the other hand, ExtremeDataHub control plane consists of Northbound-API (NB API), Southbound-API (SB API), local edge controller/manager running on edge devices, and the global controller/manager that could run on switches/controllers in case of enterprises (i.e., coordinated environment of multi ExtremeDataHub devices), or in cloud in case of uncoordinated ones. ExtremeDataHub control plane is a logically centralized entity in charge of managing and controlling ExtremeDataHub data plane through Southbound API while providing abstract network and data views through Northbound API to applications and services.

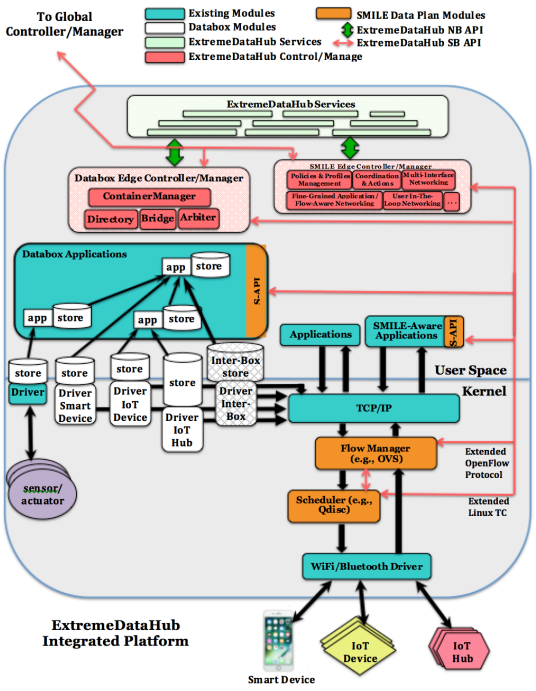

Figure 1: ExtremeDataHub architecture

We are developing ExtremeDaraHub on Linux-based small-factor hardware such as Raspberry Pi (as an example of a standalone ExtremeDataHub platform) as well as powerful hardware such as net5501-70 Board from Soekris Inc. (as an example of an integrated ExtremeDataHub platform into WiFi access points). We envision that user, enterprise, and IoT smart based environments data could be either maintained as independent individual ExtremeDataHub devices or as a networked system of multiple devices that are either geographically located within the same network or distributed across multiple network environments. In this vision, an application trying to access data should not be aware about the physical location of the data and multiple ExtremeDataHub devices should seamlessly inter-connect together to enable accessing to the required data regardless of the physical location of these data. In Figure 1, Inter-Box store and Driver Inter-Box modules are responsible in this seamless interconnection between multiple ExtremeDataHub devices.

International Collaboration:

In this project, we are collaborating with Dr. Hamed Haddadi, a senior lecturer in the Dyson School of Design Engineering at The Faculty of Engineering at Imperial College London. Dr. Haddadi is also an Academic Fellow of the Data Science Institute. His research focuses on the interaction between sensors, algorithms, and their users, data privacy and edge computing. At Imperial, he leads the Systems & Algorithms laboratory.