Towards Security in Distributed Home System

Jianxin Zhao[*], Richard Mortier[*], Hamed Haddadi[†] and Jon Crowcroft[*]
[*]The Computer Lab, Univeristy of Cambridge, UK. Email: first.last@cl.cam.ac.uk
[†]School of EECS, QMUL, UK. Email: first.last@qmul.ac.uk

***Abstract*—Today, personal data analytics and privacy face a dichotomy: application authors and service providers require scalable analytics systems, while the users and regulators increasingly demand for applications which respect the individuals' privacy. In this paper we propose to use VPN to solve new security challenges in a distributed home network system. On a prototype implementation, our initial evaluations indicate that we can feasibly provide a scalable trusted and distributed data aggregation and processing platform with acceptable overheads, while providing data privacy for users of the system.**

I. Introduction

People are increasingly surrounded by digital devices, from traditional computers to mobile phones, tablets and to numerous smart home IoT devices such as light, camera and locks. Platforms are emerging that simplify the management of these devices and use of personal data. HomeOS [1] recognizes the heterogeneity across homes in terms of devices and interconnectivity, and thus aims to bridge the gap by providing a PC-like abstraction for network devices to users and developers. Bolt [2], on the other hand, focuses on fine-grained data management to manipulate data from connected devices in home.

Besides device heterogeneity and data management, the surge in personal data generation and use also causes other challenges, one of them being tension in the collection and use of personal data, between the benefits to various analytics applications, the privacy consequences and security risks.

Our response is to provide technical means to assist the data subject in managing access to their data by others. As we have previously proposed, the Databox is an open-source personal networked device augmented by cloud-hosted services that collates, curates, and mediates access to our personal data, under the data subject's control [3]. It sits within an ecosystem of networked devices and associated services enabling individuals to manage their data, and to provide other parties with controlled access to them. Composed of a set of service instances, realised as Docker-managed containers in our current prototype, it enables Cloud-Assisted Networking through the placement of these instances in different locations, from a physical device in the subject's home, to the public cloud, to future envisioned edge-network hosting resources such as smart lampposts and cell-towers. An architectural design is shown in Fig. 1.

II. Research Question

My research focus on the security of distributed Databox system. Security is a challenging issue in Databox. First, it aims to deal with all of a user's digital devices such as laptops and mobile phones, via cable, WiFi or mobile network connections, rather than just home IoTs in a stable home network environment; Second, Databox envisions data be shared and traded among different parties. Sharing data across homes or between neighbours is a common requirement for personal data management. Trading personal data with interested organizations or companies will also be a norm for Databox users.Therefore we maintain that distribution should be a key factor in Databox, and that secure communication channel is of vital importance. Such issues are not considered in previous work. [1] only employs access control to protect user privacy, while the security feature [2] provide is to encrypt data chunks on storage. Both mechanisms are insufficient to protect user data privacy.

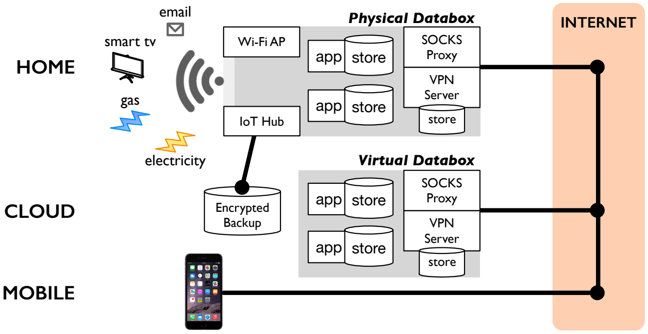

Fig. 1. Databox configured to capture an individual's online activity, in all locations, across all devices.

We propose to solve this problem by utilizing virtual private network (VPN) to securely connect different devices to create one cohesive network. So traffic of user devices are forwarded through VPN proxy to the Databox as shown in Fig 1. This method thus provide a secure encrypted connection as well as a unified interface to collect data across devices.

III. Initial Approach and Evaluation

Currently we are developing the prototype of Databox, and we would like to investigate two questions: 1) the feasibility of ARM platform as prototype machine, and 2) how does the VPN solution affect the distributed Databox system?

ARM platform has a good potential to replace x86 in cloudlet implementation. The main advantage of such platform resides in its low price and power consumption. It is commonly believed to be unsuitable to build home network systems. That's why previous work build prototypes on PC and cloud machines rather than ARM machines. If its feasibility to

TABLE I
The throughput measured by iPerf in Mbps, for different “client-server” combinations. Each result is shown in the form of *mean (standard deviation)*.

	rPi	Android	PC
Direct Conn.	25.33 (4.78)	60.9 (13.78)	24.15 (9.25)
rPi-WiFi	2.36 (0.86)	14.26 (8.49)	3.70 (0.94)
rpi-Wired	10.6 (2.38)	33.6 (7.66)	8.58 (3.57)
Desktop-Wired	15.79 (4.67)	41.37 (13.82)	11.85 (3.81)

support a Databox prototype is proved, we can easily scale up the deployment of Databox.

As a first step to investigate distribute Databox system, we explore a specific case: dividing the Databox into a Physical (local) part and a Virtual (cloud) part, i.e. a hybrid architecture for Databox. Data sources are not necessarily remote cloud servers – they may be home IoT devices or mobile phones. This begs the question as to weather different configurations of cloud and home Databoxes may introduce changes in latencies between when a datum is sampled and when it becomes available for processing.

A. System Performance

We compare Raspberry Pi 3 with a Desktop (Intel Core i7, 2.93 GHz, 8GB RAM) as the home Databox server. We use a Samsung P428 Laptop PC, another Raspberry Pi and an Nexus 5x Android phone as possible client devices. Both Databox server can be connected to a WiFi or Ethernet, and clients all connected to WiFi. We set up another machine that runs an iPerf3 service, which the client can connect to measurement its throughput. We use “ping” to estimate latency from client to this machine, and `iperf3` to measure throughput on clients. Each measurement is repeat 100 times.

In this controlled environment, we compare latency and throughput performance when the client is connected to the iPerf server directly or via Databoxes with different devices and network connections. The result are shown in Table. I. The result of latency are not shown here due to space limitation.

From initial performance tests we find that compared with Desktop being server and direct connection, client’s throughput are significantly degraded (throughput reduced by over 90%) when installed on the low power ARM-based devices, but the gap is greatly mitigated when use 1Gbps wired connection instead of WiFi on rPi. Although this is an unoptimised setup, the result shows acceptable overheads, and seems to support that the computing power difference between ARM-based devices and Desktops are not as crucial as expected.

B. Hybrid Architecture Evaluation

As previously mentioned, common Databox use cases involve data sources in the home, such as IoT devices, or mobile data sources, such as smartphones. For these tests, we measure *Time to Availability* (TTA) which denotes the time between when a datum is emitted by a source, and made available within the Databox environment. We examine how TTA changes when different configurations of Databox architecture are used, and which of these scenarios are most ideal when considering data source locus.

With a non-cloud-based device as a data source, tests are run under four possible paths the data can flow and destinations it can reach: Device to cloud Databox directly, Device to home Databox via cloud VPN, Device to home Databox directly, Device to cloud Databox via home VPN. We measure TTA under each scenario for two cases each: data source is in or outside the home. We use real data in experiments, and the source is a mobile phone streaming high frequency data – in this case accelerometer readings. Mobile phone is connected to the Internet over Home WiFi or a normal cellular network. The first case additionally acts as a close approximation of other home-based sensors and IoT devices.

Results show that when a data source is in the home, it is preferable for the data to be processed in the home, while a cloud databox shows advantages when the data source is not. We note that in all four cases the time differences of processing data on the cloud or in home are so minor that users are difficult to distinguish. Indeed, from a user experience perspective, he would be indifferent to where his data are processed for performance reasons alone.

The results support the hybrid approach and suggest that correctly managing heterogeneous resources will be a key challengein Databox.

IV. Conclusion and Future Direction

This paper briefly introduces Databox, a hybrid locally- and cloud-hosted privacy-enhance system for personal data management. By using VPN, we aim to investigate security issues in the distributed Databox system. The current prototype implementation is a first step towards realisation and evaluation of our proposed architecture. We assess the impact of security mechanisms at the user end, and evaluate a variety of architectural configurations for providing end-to-end encryption for data collection from sources such as mobile sensors and IoT devices, using VPN services on the cloud or a physical Databox at the network edge.

Further extension of this work seeking to understand the dynamics of how instances of these services could migrate on demand between cloud and home. Ephemeral caching of data in cloud stores is likely an avenue of improvement worth pursuing. The case for edge processing also worths further investigation considering the increasing data from IoT devices and home sensors. To investigate the performance of security enhancing methods in a distributed Databox networked system is another exciting explore direction.

References

[1] C. Dixon, R. Mahajan, S. Agarwal, A. Brush, B. Lee, S. Saroiu, and P. Bahl, “An operating system for the home,” in *Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation*. USENIX Association, 2012, pp. 25–25.

[2] T. Gupta, R. P. Singh, A. Phanishayee, J. Jung, and R. Mahajan, “Bolt: Data management for connected homes.” in *NSDI*, 2014, pp. 243–256.

[3] A. Chaudhry, J. Crowcroft, H. Howard, A. Madhavapeddy, R. Mortier, H. Haddadi, and D. McAuley, “Personal data: thinking inside the box,” in *Proceedings of The Fifth Decennial Aarhus Conference on Critical Alternatives*. Aarhus University Press, 2015, pp. 29–32.