2020 IGF Best Practice Forum – Data and New Technologies

IoT-based scenario case study on the collection and use of users' data and best practices

Prof. Dr. Christine Tan, *PhD CEng FIET*

China-FIOT Open Lab; Fujian Normal University

1. Background information

The Internet of Things (IoT) is becoming increasingly ubiquitous and Gartner market research has forecast around 5.8 billion IoT device endpoint connections by year 2020 [1]. In common daily scenarios at work, play and home, there are more smart devices (IoT device endpoints connected to the Internet and cloud), mostly centered around humans and making daily activities easier, less laborious and oftentimes cheaper.

Personal privacy and data security are now critical considerations as large volumes of data are now collected from IoT devices. For example, users' geographical location data tracked when using a wearable, mobile phone or even an App, can reveal rich underlying information about a user's personal preferences at the least, or in extreme cases, provide incriminating or exonerating evidence regarding a crime. Another example is healthcare data such as personal genome, where patients might want to actively decide which entity (*e.g.* hospital, insurance companies) has access to selective parts of their genomic information.

How users' data are collected and used, as well as the security protocols to manage these data, are now becoming more important. This case study will highlight current weakness in some IoT-based applications, explaining how users' data could be compromised in the system, despite conscious attempts by the data collectors to maintain data privacy. These weaknesses could be due to technical incompatibilities, or falling into a *"gray"* blind spot area that is not owned by a stakeholder *etc*. This case study also attempts to suggest some best practices in the form of practical solutions to the above mentioned weakness.

2. Case Study

2.1. Collection and use of users' data

Example 1: Centralized data conduits by providers

Today, many people trust centralized services that are backed by state-owned or authoritative institutions (*e.g.* banks, telecom operators). However, there is a tradeoff between centralization and market power -- inherent in this trust, people have also ceded their data privacy and bargaining power to these central providers [2]. In the burgeoning IoT industry, central providers *e.g.* telecom operators and local wireless access network providers, offer subscription access to wireless internet (4G, NB-IoT), as well as cloud-based management platforms.

One of the challenges inherent in the use of centralized service in the possibility of censorship or data monitoring by the service providers. Within China, applications such as Facebook, Twitter, Google cannot be accessed, partly due to IP addresses blocked by the telecom operators. In Singapore, the Infocomm Media Development Authority regulates websites that are flagged as illegal vice, deviant pornographic practices *etc.*; again, network operators are sanctioned to intercept messages to assist law enforcement, judicial or government agencies [3]. Another common practice in the IoT industry is the unnecessary routing of data uploaded by IoT devices to an additional intermediary, before the data is finally routed to the end goal server/cloud platform. This intermediary can be a middle-man platform that simply exists because users are using a service by a specific provider; an analogy to traditional systems would be the "calls dispatch centers". Once user volume is accrued, these intermediaries can become powerful inherent to their access to users' data and the analysis of big data derived thereof, and data privacy can quickly become an issue.

The fine print clauses in user agreement essentially dictate users to agree to the providers' terms if they want to continue to use the services. While user agreements are a socially responsible tool for informing users, in reality users may find it difficult to understand the legal jargon, may not have time to read through the extensively long document, or find themselves in the "all or nothing" situation where users' need to use the service override their discomfort in how their data privacy are handled.

Example 2: State organized expert database

An interesting case study is the "Special Commissioned Scientific Experts (科技特派员 Ke Ji Te Pai Yuan)" database, first initiated in 1999 by the government in Nanping city, Fujian province, P.R. China. The intention is to select and dispatch scientific experts, with the goals of using new technology to rejuvenate rural villages and alleviate poverty. Gallup World Poll conducted on citizen trust in governments reveal that confidence in national government in BRIICS countries is relatively higher than OECD countries [4]. As such, this expert database hold authority and esteem both from political and technical perspectives. As of 2016, 729,000 technical persons are listed as experts [5], deployed in 51,400 for-profit agriculture-related entities, serving 60 million farmers, helping to bootstrap 15,900 new enterprises and establish 16,000 "Scientific Experts" workstations [6].

There is a public database featuring the scientific experts' background, expertise area, phone contact number and a work diary recording technical interactions with local farmers. The database manager/provider has attempted to maintain the scientific experts' privacy by representing the phone number in the database with a series of "X"s. However, if a user tries to call the expert's phone number, the actual string of numbers (expert's actual phone number) will then appear on the mobile phone unit's local Android or Apple iOS system. This weakness in the system can be attributed to a

common problem – a technical issue of compatibility of data protocols between different systems and how this affects data hand-over between IoT devices and systems.

2.2. Current problems

In the first example, users' data privacy can be compromised if there are insufficiencies in policies to govern how centralized service providers, government agencies, regulatory bodies, and intermediaries in the data transmission chain access users' data.

In the second example, data protocols between two different systems are incompatible, resulting in insecure data handover between the Experts Database and local phone operating system. Even though the database manager/provider has taken care to anonymise the scientific expert's phone number, this phone number is still exposed during a call. The database manager-provider and phone operating system companies have the issue of data not in sync.

3. Best Practices

Best practices in the form of practical solutions, both currently deployed and under research, are suggested.

Firstly, there is a need for social responsibility of data collectors to ensure users' data privacy and data usage within permission limits. These centralized providers and/or intermediaries should offer open channels of communication and feedback, actively listen to users and respond to their feedback. Furthermore, the users' agreement and policies stating how data will be used, should be open for public inspection. The feedback mechanism need to work both ways for a successful outcome; users should also play an active role in providing feedback and defining who access which parts of their data.

Secondly, there is a need towards more unified data protocol standards, especially in the highly fragmented IoT industry. Vendors of IoT devices, systems and databases each develop their own data transfer protocols, so it is common to experience incompatibility of data transferred from one system to another. Ideally a neutral, independent party like FIOT Open Lab, can lead industry efforts to standardize IoT security and data protocols, however in reality, it is often difficult to achieve consensus amongst the many stakeholders involved.

Finally, there are emerging technological solutions such as blockchain that can increase trust between users and the service providers, as well as allowing users to pre-determine permissions and decide who gets to access which parts of their data. An example of such a blockchain is zkLedger [7], a new blockchain protocol developed at MIT that keeps contents of data transactions private, while allowing a third party to access certain parts of the data or query the system, without fully revealing contents of data transactions.

Society today is becoming more digital as people consume more information and video streaming, as well as exchange more data, thanks to the rapid progress in 5G telecommunication, electronic products and IoT technologies. There is an urgent need to protect data privacy, come up with policies that regulate how data is used and managed, as well as novel technologies to solve weakness that are present in the current system.

References

[1] https://www.gartner.com/en/newsroom/press-releases/2019-08-29-gartner-says-5-8-billion-enterprise-and-automotive-io

[2] S. Athey, C. Catalini, C. Tucker, *"The Digital Privacy Paradox: Small Money, Small Costs, Small Talk",* 2018.

[3] https://thelawreviews.co.uk/edition/the-technology-media-and-telecommunications-review-edition-10/1211270/singapore

[4] OECD (2013), "Trust in government, policy effectiveness and the governance agenda", in *Government at a Glance 2013*, OECD Publishing, Paris. DOI: https://doi.org/10.1787/gov_glance-2013-6-en

[5]http://www.scio.gov.cn/m/xwfbh/xwbfbh/wqfbh/2014/20140213/zy30387/Document/1362969/1362969.htm

[6] http://finance.china.com.cn/roll/20160301/3608406.shtml

[7]https://static1.squarespace.com/static/59aae5e9a803bb10bedeb03e/t/5aa1b35ce4966bd538d3f1d2/1520546653653/zkledger.pdf