Journal of Network and Computer Applications 64 (2016) 137–153

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

When things matter: A survey on data-centric internet of things

Yongrui Qin [a,*], Quan Z. Sheng [a], Nickolas J.G. Falkner [a], Schahram Dustdar [b], Hua Wang [c], Athanasios V. Vasilakos [d]

[a] School of Computer Science, The University of Adelaide, Adelaide SA 5005, Australia
[b] Institute of Information Systems, Vienna University of Technology, Austria
[c] Center for Applied Informatics, Victoria University, Melbourne, Victoria 3122, Australia
[d] Department of Electrical and Computer Engineering, National Technical University of Athens, Greece

ARTICLE INFO

Article history:
Received 18 August 2014
Received in revised form
22 June 2015
Accepted 28 December 2015
Available online 11 February 2016

Keywords:
Internet of Things
Data management
RFID systems
Wireless sensor networks

ABSTRACT

With the recent advances in radio-frequency identification (RFID), low-cost wireless sensor devices, and Web technologies, the Internet of Things (IoT) approach has gained momentum in connecting everyday objects to the Internet and facilitating machine-to-human and machine-to-machine communication with the physical world. IoT offers the capability to connect and integrate both digital and physical entities, enabling a whole new class of applications and services, but several significant challenges need to be addressed before these applications and services can be fully realized. A fundamental challenge centers around managing IoT data, typically produced in dynamic and volatile environments, which is not only extremely large in scale and volume, but also noisy and continuous. This paper reviews the main techniques and state-of-the-art research efforts in IoT from data-centric perspectives, including data stream processing, data storage models, complex event processing, and searching in IoT. Open research issues for IoT data management are also discussed.

1. Introduction

The Internet is a global system of networks that interconnect computers using the standard Internet protocol suite. It has significant impact on the world as it can serve billions of users worldwide. Millions of private, public, academic, business, and government networks, of local to global scope, all contribute to the formation of the Internet. The traditional Internet has a focus on computers and can be called the Internet of Computers. In contrast, evolving from the Internet of Computers, the Internet of Things (IoT) emphasizes things rather than computers (Ashton, 2009). It aims to connect everyday objects, such as coats, shoes, watches, ovens, washing machines, bikes, cars, even humans, plants, animals, and changing environments, to the Internet to enable communication/interactions between these objects. The ultimate goal of IoT is to enable computers to see, hear and sense the real world. It is predicted by Ericsson that the number of Internet-connected things will reach 50 billion by 2020. Electronic devices and systems exist around us providing different services to the people in different situations: at home, at work, in their office, or driving a car on the street. IoT also enables the close relationship between human and opportunistic connection of smart things (Guo et al., 2013).

There are several definitions or visions of IoT from different perspectives. From the viewpoint of services provided by things, IoT means “a world where things can automatically communicate to computers and each other providing services to the benefit of the human kind” (CASAGRAS, 2000). From the viewpoint of connectivity, IoT means “from anytime, anyplace connectivity for anyone, we will now have connectivity for anything” (ITU, 2005). From the viewpoint of communication, IoT refers to “a world-wide network of interconnected objects uniquely addressable, based on standard communication protocols” (INFSO, 2008). Finally, from the viewpoint of networking, IoT is the Internet evolved “from a network of interconnected computers to a network of interconnected objects” (European Commission, 2009).

We focus on our study of the Internet of Things from a *data perspective*. As shown in Fig. 1, data is processed differently in the Internet of Things and traditional Internet environments (i.e., Internet of Computers). In the Internet of Computers, both main data producers and consumers are human beings. However, in the Internet of Things, the main actors become *things*, which means things are the majority of data producers and consumers. Therefore, we give our definition of the Internet of Things as follows:

“*In the context of the Internet, addressable and interconnected things, instead of humans, act as the main data producers, as well as the main data consumers. Computers will be able to learn and gain information and knowledge to solve real world problems directly with the data fed from things. As an ultimate goal, computers enabled by the Internet of Things technologies will be able to sense and react to the real world for humans.*”

* Corresponding author.
E-mail address: yongrui.qin@gmail.com (Qin).

http://dx.doi.org/10.1016/j.jnca.2015.12.016
1084-8045/

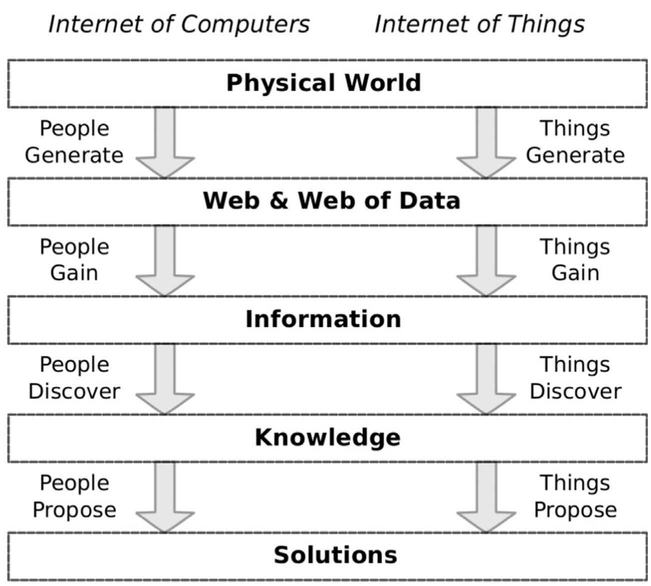

Fig. 1. Internet of Computers vs. Internet of Things.

As of 2012, 2.5 quintillion (2.5×10^{18}) bytes of data are created daily.[1] In IoT, connecting all of the things that people care about in the world becomes possible. All these things would be able to produce much more data than nowadays. The volumes of data are vast, the generation speed of data is fast and the data/information space is global (James et al., 2009). Indeed, IoT is one of the major driving forces for *big data analytics*. Given the scale of IoT, topics such as storage, distributed processing, real-time data stream analytics, and event processing are all critical, and we may need to revisit these areas to improve upon existing technologies for applications of this scale.

In this paper, we systematically investigate the key technologies related to the development of IoT and its applications, particularly from a data-centric perspective. The aim of this work is to provide a better understanding of the current research activities and issues. Fig. 2 shows the roadmap of this paper. As can be seen from the figure, we review and compare technologies including data streams, data storage models, searching, and event processing technologies, which play a vital role in enabling the vision of IoT. We also describe some relevant applications from several representative areas. Although some reviews about IoT have been conducted recently (e.g., Atzori et al., 2010; Zeng et al., 2011; An et al., 2013; Perera et al., 2013; Li et al., 2016; Yan et al., 2014), they focus on high level general issues and are mostly fragmented. In addition, these articles do not specifically cover techniques on data processing and management, which is fundamentally critical to fully embrace IoT. To the best of our knowledge, this is the first article that studies and discusses state-of-the-art techniques of IoT from the data-centric perspective.

The remainder of the article is organized as follows. Section 2 identifies an IoT data taxonomy. Section 3 reviews the data streaming techniques and Section 4 focuses on the data models and storage technologies for IoT. Search and event processing technologies are discussed in Sections 5 and 6, respectively. In Section 7, some typical ongoing and/or potential IoT applications where data techniques for IoT can bring significant changes are described. Finally, Section 8 highlights some research open issues on IoT from the data perspective and Section 9 offers some concluding remarks.

2. IoT data taxonomy

In this section, we identify the intrinsic characteristics of IoT data and classify them into three categories, including *Data Generation, Data Quality*, and *Data Interoperability*. We also identify specific characteristics of each category, and the overall IoT data taxonomy is shown in Fig. 3.

2.1. Data generation

- *Velocity*: In IoT, data can be generated at different rates. For example, for GPS-enabled moving vehicles in road networks, the GPS signal sampling frequency could be every few seconds, every few minutes, or even every half an hour. But some sensors can scan at a rate up to 1,000,000 sensing elements per second.[2] On one hand, it is challenging to handle very high sampling rates, which require efficient processing of the fast generated data. On the other hand, it is also challenging to deal with low sampling rates, due to the fact that some important information may be lost for data processing and decision making.
- *Scalability*: Since things are able to continuously generate data together with the foreseeable excessively large number of things, the IoT data is expected to be at an extremely large scale. It is easy to image that, in IoT data processing systems, scalability will be a long standing issue, aligning with the current Big Data trend.
- *Dynamics*: There are many dynamic elements within IoT data. Firstly, many things are mobile, which will lead to different locations at different times. Since they will move to different environments, the sensing results of things will also be changing to reflect the real world. Secondly, many things are fragile. This means the generated data will change overtime due to the failure of things. Thirdly, the connections between things could be intermittent. This also creates dynamics in any IoT data processing system.
- *Heterogeneity*: There will be many kinds of things potentially connecting to the Internet in the future, ranging from cars, robots, fridges, mobile phones, to shoes, plants, watches, and so on. These kinds of things will generate data in different formats using different vocabularies. In addition, there will be assorted IoT data processing systems, which will also provide data in customized formats to tailor different data needs.

2.2. Data quality

- *Uncertainty*: In IoT, uncertainty may come from different sources. In RFID data, the uncertainty can refer to missing readings, readings of non-existing IDs, etc. In wireless sensor networks, uncertainty can refer to sensing precision (the degree of reproducibility of a measurement), or accuracy (the maximum difference that will exist between the actual value and the indicated value), etc.
- *Redundancy*: Redundancy can also be easily observable in IoT. For example, in RFID data, the same tags can be read multiple times at the same location (because multiple RFID readers exist at the spot or tags are read multiple times at in the same spot) or at different locations. In wireless sensor networks, a group of sensors of the same type may also be deployed in a nearby area, which can produce similar sensing results of that area. For the same sensor, due to the possible high sampling rates, redundant sensing data can be produced.
- *Ambiguity*: Dealing with a large amount of ambiguity in IoT data is inevitable. The data produced by assorted things can be interpreted in different ways due to different data needs from different things or other data consumers. Such data can also be useful and important to

[1] http://www-01.ibm.com/software/data/bigdata/

[2] https://www.tekscan.com/support/faqs/what-are-sensors-sampling-rates