Ubiquity, an ACM publication
October 2015

Ubiquity Symposium

The Internet of Things

Discovery in the Internet of Things

by Arkady Zaslavsky and Prem Prakash Jayaraman

Editor's Introduction

How to find a "thing" in the Internet of Things (IoT) haystack? The answer to this question will be the key challenge that IoT users and developers are facing now and will face in the future. Current models for IoT are focused heavily on developing vertical solutions limited by hardware and software platforms and support. With the estimated explosion of IoT in the coming years as predicted by Cisco, IBM and Gartner, there is a need to rethink how IoT can deliver value to the end-user. A paradigm shift is required in the underlying fundamentals of current IoT developments to enable a wider notion of "thing" discovery as well as discovery of relevant data and context on the IoT. Discovery will allow users to build IoT apps, services and applications using "smart things" without the need for a priori knowledge of things. In this article, we look at the current state of IoT and argue for paradigm shift addressing why and how discovery can make a significant impact for the future of IoT and moreover, become a necessary component for IoT success story.

Ubiquity Symposium

The Technological Singularity

Discovery in the Internet of Things

by Arkady Zaslavsky and Prem Prakash Jayaraman

A recent forecast made by IDC projects the Internet of Things (IoT) and the associated ecosystem to be a $1.7 trillion market by 2020, which will include 212 billion connected things. A recent Gartner hype cycle report estimates IoT to be at the peak of inflated expectations [1]. The IoT will fuel a paradigm shift of a "truly connected" world in which everyday objects become inter-connected and smart with the ability to communicate many different types of information with one another as well as with human users. Figure 1, presents a graphical forecast of IoT explosion over the coming years as estimated by Cisco.

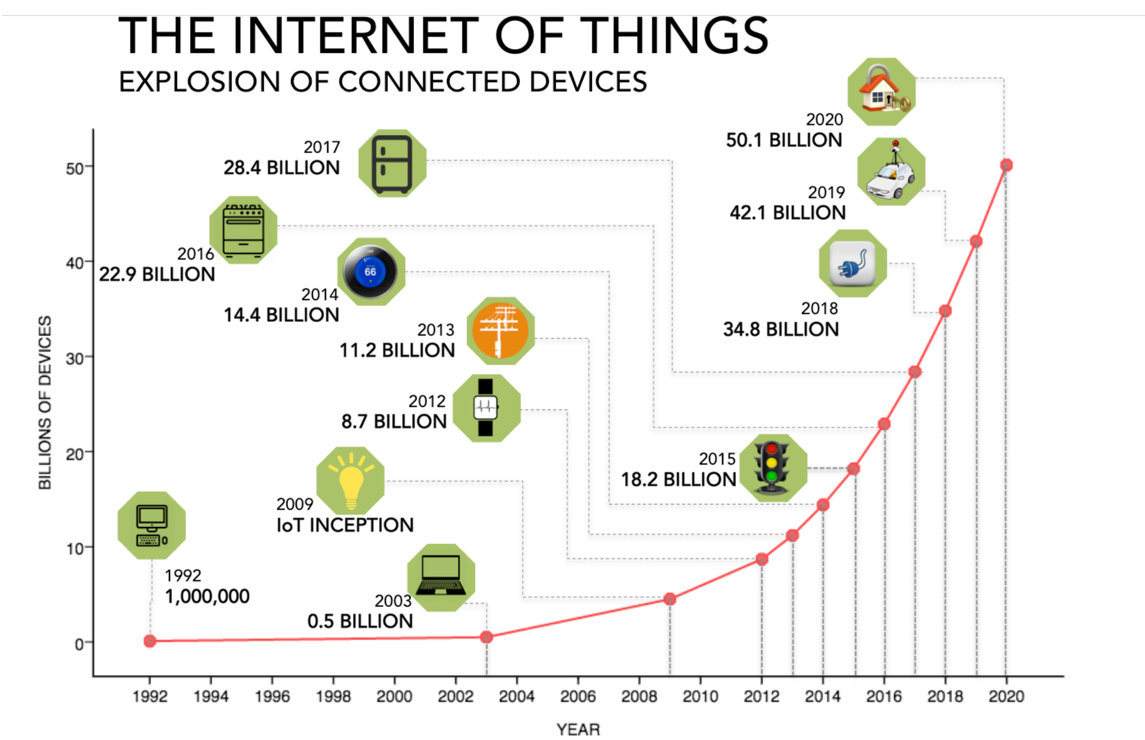

Figure 1. The explosion of Internet of Things

The term "Internet of Things" collectively describes technologies and research disciplines that enable the Internet to reach out into the real world of physical objects. Technologies like RFID, short-range wireless communications, real-time localization, and sensor networks are becoming increasingly pervasive making the IoT a reality. Sometimes, IoT is synonymized with cyber-physical systems (CPS). It is envisioned the IoT paradigm will open up the possibility to create novel value-added services across areas of science, technology, business, economy, and so on by dynamically combining different types of capabilities (e.g. sensing, communication, information processing, and actuation on physical resources) [2].

The Current Landscape of IoT

The IoT is expected to contribute towards the ambitious vision of the creation of a large-scale, "smart" interconnected world of machines, devices, sensors, actuators and systems including human users. Figure 2 depicts the current IoT landscape and the various areas of applications including industrial, consumer, and automotive. The IoT data, however, is not limited to sensors and machines but data from social networks, the web, and other user submitted physical observations and measurements [3]. The current IoT landscape comprises many large players such as Intel, EMC, Siemens, Microsoft, and Freescale to name a few. In the IoT, data from real or virtual world things will be available globally and in vast amounts to be shared among applications and devices for event detection, context and situational awareness based decision making, enhanced service creation, and driving event-based actuation without human intervention. Let us consider a futuristic smart home scenario in a "smart" interconnected IoT world. The user's car can provide information about when the user departs from work, GPS can verify the destination is home, and using traffic data, the estimated arrival time can be determined. Using this data a smart heating, ventilation and air conditioning system (HVAC) can itself turn on and ensure a comfortable temperature is maintained in the house before the user's arrival. Such a system has to be reactive, efficient, and effective as it will have to continuously respond to changes in the user's situation e.g. delay due to a traffic jam. However, such a system can also be proactive where proactive behavior is based on predicted situations evaluated with some level of confidence and probability.

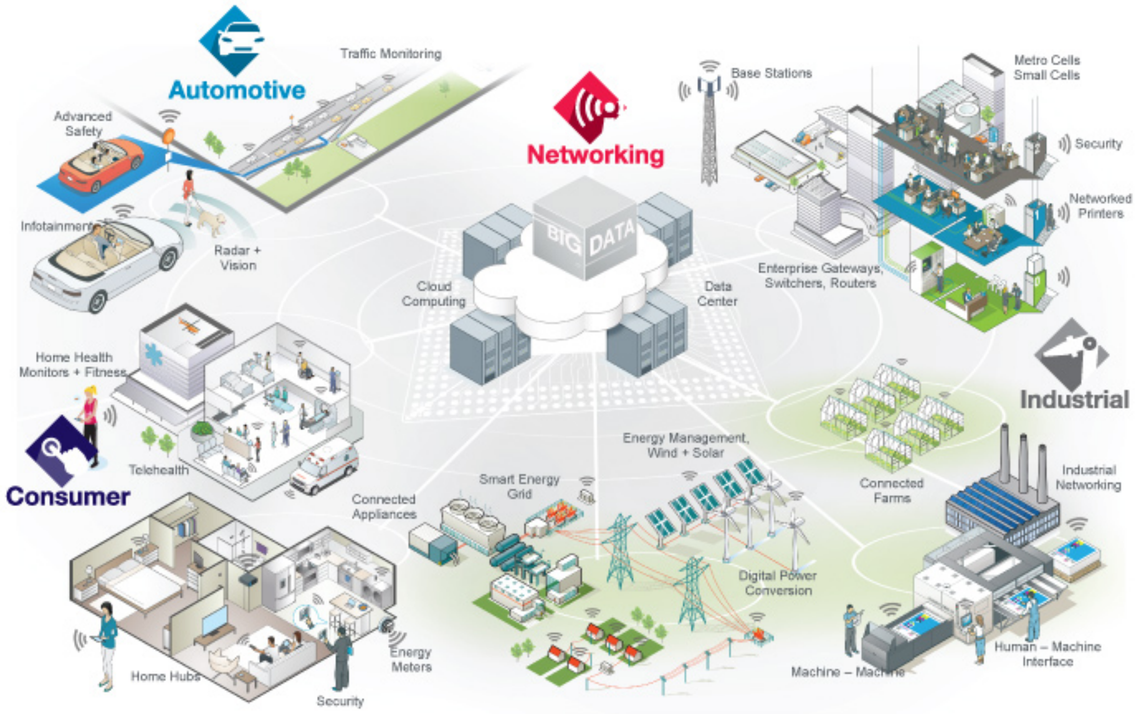

Figure 2. Internet of Things Landscape © Freescale (Reprinted with permission)

Discovery: What is it?

> *"The acquisition of knowledge is always of use to the intellect, because it may thus drive out useless things and retain the good. For nothing can be loved or hated unless it is first known."*—Leonardo da Vinci

Discovery is a mechanism that will enable application to access the IoT data without the need to know the actual source of data, sensor description, or location. Librarians use discovery to describe search tools, which are aimed at a category of "learning" users (e.g. students). Discovery or data discovery is used in relation to big data applications mainly to describe "visual analytics" tools. Discovery is also used in science, for example drug discovery to describe activities leading to the creation of new knowledge.

Defining discovery is a challenging task because it corresponds to activities that are specific to data providers, e.g. the curation of tasks required before the publication of a specific datasets, and other activities that are specific to end publishers or brokers accessing and integrating multiple datasets to support data linking and context-driven search. The discovery process can be defined as two successive loops:

- **Foraging loop.** Data sources are identified and assessed, where the relevant data is extracted and formatted into consumable form.
- **Sense-making loop.** The extracted data is analyzed and exploited to provide answers around a specific problem

The challenge is then to develop a framework (or architecture) to provide complete capabilities, which works for proponents of big data and the IoT.

Why Discovery?

An inherent characteristic of the IoT is "heterogeneity" introduced by a plethora of things with different data communication capabilities (protocols and hardware, data rates, reliability, etc.); computational, storage and energy capabilities; diversity in the types and formats of data (audio, video, text, numeric, and streams); and IoT standards (device standards, standards to represent data, IEEE projects on IoT standards, ITU and ISO IoT standards, etc.). The diversity in things and the data produced by them pose significant challenges in fulfilling the ambitious dream of a truly interconnected smart world of things.

Further, it is expected the IoT will be a major source of big data driven by its velocity, variety, value and volume. The diverse IoT data will be in high demand by business and end-user applications and hence will have to be stored in widely distributed, heterogeneous information systems to ensure global availability. However, retrieving the data from these heterogeneous data stores is a non-trivial task without a common machine-readable data representation framework. Moreover, when dealing with large volumes of distributed and heterogeneous data, issues related to interoperability will need to be addressed. It is widely recognized that efficient mechanisms for discovering available resources and capabilities in the IoT is essential.

Finally, current IoT solution stacks focus on the development of innovative low-footprint hardware solutions that are integrated into vertical software middleware silos [4]. Data by itself in silos is of limited value unless analyzed with other relevant data and context. With the current technologies, we can create haystacks of data. But the real challenge here is how to find the needle across zettabytes of multiple haystacks. This problem is all the more relevant when heterogeneity and complexity in IoT data makes it hard to describe the needle precisely. This is where discovery combined with analytics in the IoT can delivery unknown insights into data produced by things. Figure 3 presents an overview of why discovery is important in the IoT context. As depicted in the figure, on the left, the current approach creates IoT data silos by tightly coupling application with specific sensors (e.g. vendor specific solutions); while on the right is our discovery-based IoT vision where applications and sensors are loosely coupled, allowing interoperability, discovery and re-use/re-purposing of IoT data.

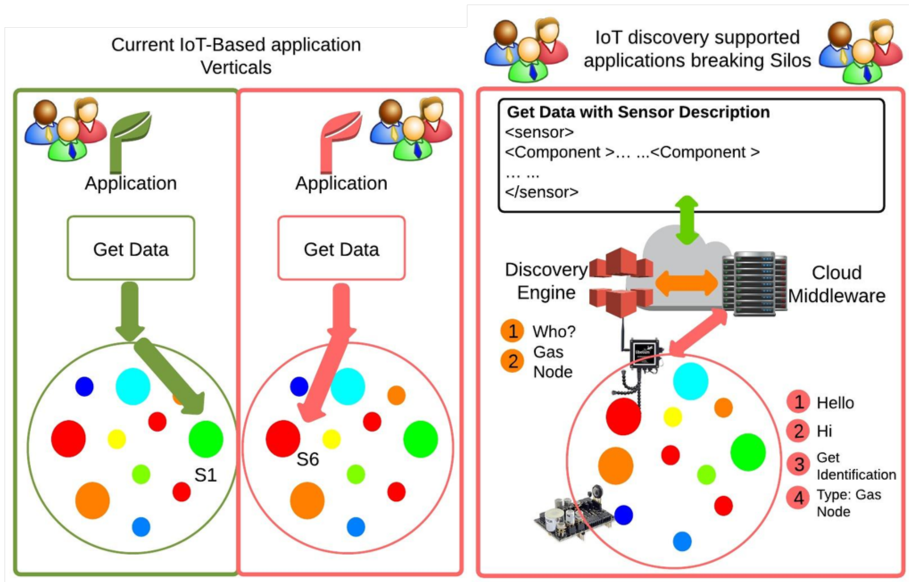

Figure 3. Internet of Things–Discovery

Discovery in the Internet of Things: Overview

As shown in Figure 4, a user trying to discover data about a real-world thing has multiple pathways to try to find it: Exploring structured data or textual data first—or exploring them simultaneously—or focusing on the identification of live data resources (there are not many textual resources that provide insights on such data except maybe Twitter or an equivalent). The picture also highlights the challenges of shifting focus from the digital world to the real world if indeed the object of discovery is a real-world entity or phenomenon. Sensors and/or actuators (IoT things) play an important proxy role for this latter phase. It is important to note the difference between use cases where things are named before they are sensed and use cases where things are sensed before they named (identified). With disruptive technology changes such as IoT and big data, we need to support new data/sensor discovery to overcome a number of issues preventing seamless access and reuse of data, and more specifically "live" data, coming from things.

Information discovery is different for web-based information locators primarily designed for text-based data [5]. By using information discovery approaches, such as the [Google Knowledge](#)

Graph, greater insights into data can be obtained. For example, Knowledge Graph attempts to understand the information (structured, semi-structured and unstructured) on the Web by dynamically connecting facts about people, places, and things (referred to as entities). The Google Knowledge Graph moves away from standard webpage search to exploratory mode by providing relations and answers to questions the user never intended to ask. The Knowledge Graph is a semantic search approach that uses a standard ontology to infer facts about entities. Recent work published by Google in 2014 proposed the Google Knowledge Vault [6], which uses a probabilistic approach to build a probabilistic knowledge graph that can automatically build knowledge about entities by attaching a certain level of confidence to each relation. This allows the knowledge vault to distinguish what is known with high confidence and what is uncertain.

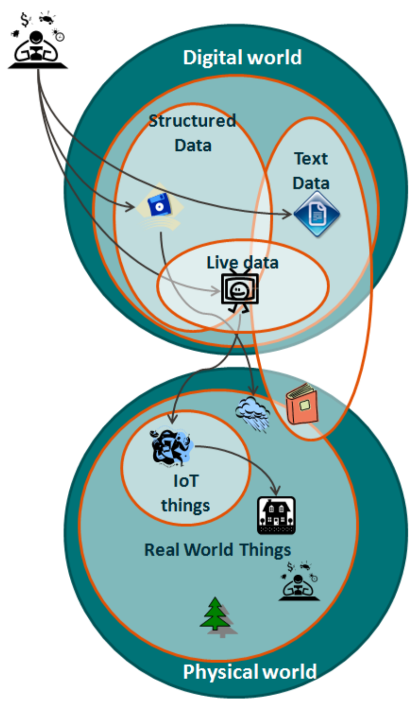

Figure 4. Discovery Loop in the IoT World

Applying approaches like knowledge vault to IoT is not a trivial task due to the big data complexities, such as volume, variety, and velocity coupled with constant changes in relationships among entities. This will require far more sophisticated approaches and techniques that can manage the volume of data produced by IoT. For example, Wörner et. al. have explored the use of data from an off-the-shelf weather station to determine room occupancy [7]. This is a relevant example of how discovery in the IoT can help extract new knowledge from live data sources, which are otherwise perceived as ordinary data sources connected to a specific application delivering outcomes as defined by the application (e.g. live weather reports and alerts). Further, by fusing data from things and building relationships among them dynamically using a probabilistic semantic approach can facilitate the inference of completely new knowledge. Building a knowledge graph of things around entities can answer the obvious questions, but also extract and discover new knowledge that the user never intended to find.

Final Thoughts

More recent work on discovery of things and IoT data is restricted to queries that are based on location, time, and type of measurement with little consideration to the entity of interest. Our vision for discovery in the IoT is to answer knowledge-based queries. Consider the scenario of a scientist who is researching the effect of locust spread on plant varieties. This data is collected via a wireless sensor network and is available with basic meta-data descriptions. The query we like to address is how this data can be repurposed for use by an entomologist, who is studying the behavior of locust and looks for ways to stop the spread. This type of knowledge discovery will require novel methods that will go beyond the type of sensor search but will also incorporate a range of reasoning techniques that can link datasets across domains. These reasoning techniques can be probabilistic, deterministic, or semantically driven. The challenge is to develop a suit of these techniques that does not suffer from performance bottlenecks (response time, processing, or reasoning) given the estimated prediction of things by 2020. Our vision is close to the vision of open data, but advances one-step further allowing application users to discover and orchestrate services over open data. Data/service/thing discovery is the step forward but the key is information, relationship extraction, and knowledge discovery.

About the Authors

Arkady Zaslavsky is a senior principal research scientist with CSIRO Data61, CSIRO, Australia. He leads strategic research projects in the Internet of Things science area. He currently holds the titles of a research professor at LTU (Sweden), adjunct professor at UNSW (Sydney), adjunct professor at La Trobe University (Melbourne), visiting professor at St. Petersburg University of ITMO. Before coming to CSIRO in July 2011, he held a position of a Chaired Professor in Pervasive and Mobile Computing at Luleå University of Technology, Sweden where he was involved in a number of European research projects, collaborative projects with Ericsson Research, PhD supervision and postgraduate education. Dr. Zaslavsky has published more than

350 research publications throughout his professional career and supervised to completion more than 30 Ph.D. students. Dr. Zaslavsky is a senior member of ACM and a senior member of IEEE Computer and Communication Societies.

Prem Prakash Jayaraman is currently a research fellow at RMIT University, Melbourne. His research areas of interest include, Internet of Things, cloud computing, mobile computing, sensor network middleware and semantic internet of things. Dr. Jayaraman is one of the key contributors of the Open Source Internet of Things project (OpenIoT) that has won the prestigious Black Duck Rookie of the Year Award in 2013. He has been the recipient of several awards including hackathon challenges at the Fourth International Conference on IoT (2014) at MIT Media Lab, Cambridge, MA and IoT Week 2014 in London and best paper award at IEA/AIE-–2010. He was a postdoctoral research fellow at CSIRO Digital Productivity Flagship, Australia from 2012 to 2015. Prior to that, he worked as a research fellow and lecturer at the Centre for Distributed Systems and Software Engineering, Monash University, Melbourne, Australia. He has served as a program committee member for the Hawaii International Conference on System Sciences and Mobile Data Management Conferences and is a reviewer of many distributed systems and software engineering journals, including Elsevier's *Future Generation Computer Systems*, John Wiley & Sons' *Concurrency and Computation: Practice and Experience*, *World Wide Web Journal*, and *IEEE Transactions on Cloud Computing*.

References

[1] Butler, B. Gartner: Internet of Things has reached hype peak. Network World. Aug. 13, 2014. Accessed Sept. 12, 2014.

[2] Paganelli, F. and Parlanti, D. A DHT-Based Discovery Service for the Internet of Things. *Journal of Computer Networks and Communications.* Article ID 107041 (2012).

[3] Sheth, A., Anantharam, P., and Henson, C. Physical-Cyber-Social Computing: An early 21st century approach. *IEEE Intelligent Systems* 28, 1 (2013), 79–82.

[4] Zorzi, M., Gluhak, A., Lange, S., and Bassi, A. From Today's INTRAnet of Things to a Future INTERnet of Things: A wireless- and mobility-related view. *IEEE Wireless Communications* 17, 6 (December 2010), 44–51.

[5] Barnaghi, P. Discovering Things and Things' data/services. Presentation. IoT Week London, UK. June 16–20 2014.

[6] Dong, X., Gabrilovich, E., Heitz, G., Horn, W., Lao, N., Murphy, K., Strohmann, T., Sun, S., and Zhang, W. Knowledge Vault: A web-scale approach to probabilistic knowledge fusion. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (KDD '14). ACM, New York, 2014, 601-610.

[7] Wörner, D., von Bomhard, T, Röschlin, M., and Wortmann, F. Look Twice: Uncover hidden information in room climate sensor data. Fourth International Conference on Internet of Things. MIT Media Lab. Cambridge, MA, October 6-8, 2014.

[8] Barnaghi. P, Wang. W, Henson. C, and Taylor. K. Semantics for the Internet of Things: Early progress and back to the future. *Int. J. Semant. Web Inf. Syst.* 8, 1 (January 2012), 1-21.

[9] Kondepu, K., Restuccia, F., Anastasi, G., and Conti, M. A Hybrid and Flexible Discovery Algorithm for Wireless Sensor Networks with Mobile Elements. 2012 IEEE Symposium on Computers and Communications (ISCC). July 1–4, 2012. IEEE.

DOI: 10.1145/2822529

