Milan Milenkovic, Intel Corporation, Internet of Things Group

Towards a Case for Interoperable Sensor Data and Meta-data Formats, Naming and Taxonomy/Ontology

Abstract

While much attention in the Internet of things/web of things (IoT/WoT) community has been focused on designing sensing systems for dedicated infrastructure or for network of standalone sensors, enabling sensing capability that spans across domains and across devices has not been seriously addressed to date. The challenge for the community is to devise standards and practices that enable integration of data from sensors across devices, users, and domains to enable new types of applications and services that facilitate much more comprehensive understanding and analyses of the world around us and ultimately improve the quality of life.

Useful elaborations of such systems can provide significant new business opportunities for services, big data, and analytics. In order to fulfil that promise, IoT/WoT systems need to be designed to support some level of processing commonality by defining interoperable sensor data and meta-data formats, naming, taxonomy and possibly ontology. This paper sketches some use cases that motivates this need and outlines the initial requirements.

Introduction

A simple but powerful view of Internet (web) of things is as a sensor-enhanced Internet, i.e. sensors attached to the Internet, directly or via intermediaries, with the ability to source data and, where appropriate, to provide actuation and possibly physical impact on the real world. Sensors usually convey information about real-world phenomena, widely ranging from direct measurements, e.g. temperature, to user observations, such as river is overflowing.

Connecting sensors adds real-world data, and optionally awareness, to the Internet. This deceptively simple addition is a transformational change, it basically bridges the gap between physical and virtual/cyber worlds that has persisted since the invention of modern computing. In effect, IoT is Internet with all of its features and capabilities with the real-world dimension (and interface) added to it. As has been observed, with IoT Internet becomes a web of people, information, services and things. This view of IoT/WoT is somewhat different from M2M systems which focus mostly on machines talking to each other over the available connections, including the Internet. Internet of everything is more inclusive in the sense that it assumes that everything is connected to everything else using Internet fabric and protocols, and thus has the potential to engage in interactions and to provide a plethora of exciting new uses and services limited only by creativity and security/privacy restrictions.

Our definition of sensors is very broad - it includes not only all types of hardware sensors and wireless networks, but also software sensors, sensing services, and people. Software sensors are usually software agents that can capture and report on some real-world condition of interest, such as user presence detected via key clicks or mouse movements. Sensing service refers to data provided by an external source with programmatic interfaces, e.g. a localized weather reporting service. People as sensors refers to users providing direct input - say on their observations, comfort or system adjustment preferences - via dedicated end-user interfaces or social networks.

Sensors are appearing in a variety of forms and settings. Many early projects, mostly academic, started with wireless sensor networks (WSNs) that were deployed "in the wild" in agriculture, for sensing bridge vibrations, and in smart cities to name a few. In addition, there are probably billions of sensors in legacy systems such as:

industrial automation, energy and health systems, building-management systems (BMS), and infrastructure. Most of their data are locked in proprietary systems, such as SCADA and BMS, but those are increasingly becoming interfaced to the Internet for at least restricted forms of sharing and remote control. Probably, the strongest numerical contributor to the installed base of Internet-connected sensors are the smart phones - more than a billion of them are already in use and a recent model contains more than ten program-accessible and thus Internet connectable sensors. And many newcomers are also showing up on the scene, including automotive Internet-connected IVI systems and personal health and fitness sensors. Hence, sensors on the Internet are not just a thing of the future - the currently installed base already numbers in the billions and growth projections are meteoric.

Opportunity

At present, much of this sensor data is of limited value as it is locked in proprietary systems and in vertical, usage- and device-specific, applications. A great opportunity lies in being able to create horizontal services and applications that make use of aggregations of sensor data across devices and domains. Such services can focus more on the needs and interests of users, e.g. as data from devices belonging to a user or of interest to the user in a particular context, surrounding, or at a given point in time. A simple early example is being able to discover sensors of interest in user's proximity, such as discovering nearby clusters of air-quality sensors in a smart city and visualizing air quality to the user on her personal smart device "right now and where I am". A car can register vibration when going over a pothole and transmit its related GPS coordinates to inform the city's maintenance department - crowdsourcing that is timely, accurate and saves labor costs. Another example would be to aggregate data from user's energy and ambient sensors at home and in the office and from the mode of transportation and used to create a personal energy or carbon footprint tracker. Such data could be collected for communities of users to provide comparisons with relevant averages of interest, such as other users or homes, and to allow target tracking and competitions to achieve individual or group goals. Examples are simple, but they highlight potential benefits of linking sensing with services across domains, such as individuals, buildings, neighborhoods, and cities.

Towards Requirements and Approach

Useful elaborations of such systems provide significant new business opportunities for services, big data, and analytics. In order to fulfil that promise, IoT systems need to be designed to support data interoperability and at least some level of commonality in taxonomies, ontologies, naming, and meta-data assignment and processing.

As an illustration, consider a scenario of creating an application that tracks a person's carbon footprint. It could start by aggregating data on that person's energy usage at the office, home and commute, possibly offset by tracking energy-saving activities, such as walking or biking to work. Later on, tracking of carbon impact of long-distance travel data could be added to complete the picture. And to make it more valuable, the system could provide relevant averages across aggregations of interest – such as footprints of people in the same company, neighborhood, city, country, and ultimately the world. This could be used for comparison purposes to train user's intuition and possibly to enable competitions or tracking of group energy saving goals in social circles of interest.

Implementing such an application would require the system to be capable to access data – directly or via protocol translators - from different domains, such as a legacy building-management system (BMS) in the office, energy home automation or smart meter at home, and vehicle system for travel distance and energy consumption tracking. In addition, it would require meta-data and sensor-user associations or attributions for data personalization, i.e. ability to retrieve user specific data on energy usage across different domains – such as home, office and car - and aggregation to provide averages across groups of interest.

In reasonably common settings, this application would use data from a variety of sensors attached to legacy systems, like BMS, IoT-aware instrumented systems like home automation and vehicles, and some personal or user-targeted sensors like office power-strip energy meter and wearable activity tracker. With today's prevailing state of affairs, this would be prohibitively costly due to expensive custom coding or even impossible as sensor data are locked in fragmented and often proprietary vertical silos, including – in this example: BMS, vehicle information system, home automation, energy meters, wearable fitness/activity tracker. A streamlined, web-style approach would be to have interoperable data and meta-data definitions that would acquire sensor data and allow aggregations and processing of interest.

IoT/WoT community should work on a minimal viable set of interoperable data and meta-data formats that work across devices and domains, preferably augmented with naming and taxonomy systems that would allow interested applications to discover and access sensors of interest, be able to interpret their observations, and perform actuations where appropriate. This would need to include basics and context such as: sensor ID/name, observation/reading, engineering units, time of acquisition, location. The list of items and meta-data of interest can quickly grow to include: sensor type, location, frequency of reporting, mobile or static location, owner, domain, associations, access rights, privacy policy and restrictions, accuracy, calibration, manufacturer, model number, and others. The challenge is to devise a coordinated naming, taxonomy/ontology and meta-data system that combine together to give minimal useful information in each observation, and allows the rest of the information of interest to be obtained by querying the sensor node or cloud data structures, as appropriate.

Depending on a system hierarchy and function placement, copies or fragments of data and meta-data may be processed and stored at the edge near the point of acquisition, at a local aggregation point – such as a gateway, and/or in a cloud. Meta-data typically change at a different, and generally slower, rate than sensor data which may favor separation of their processing and storage paths. Regardless of implementation, services need to be able to query sensor nodes and/or the cloud to obtain real-time readings/observations and to query historical sensor values, say by name and time. The system also needs to be bidirectional, so that authorized entities can carry out actuation actions – such as opening or closing of a valve - for direct impact on the physical world as and when appropriate. For aggregations and collections of data across devices and domains, it is highly desirable to be able to support searches of sensor data and meta-data by combinations of attributes (meta-data), such as sensors in my proximity, in an area (e.g. temperature sensors in the SW section on the third floor of this building), by location, by sensor type, by values/patterns/trends, by group/domain, etc.

At the outset, this is a very complex problem and it is tempting to stray into cumbersome premature generalizations and formalisms that may hinder adoption and implementation. Given that the whole area is quite fluid and fast moving, probably the best and the fastest way forward is for the IoT/WoT community to start by defining a minimal usable subset of guidelines and specifications with room for subsequent growth and expansion as our experience with building and operating those systems evolves, much like the evolution and success of the worldwide web. As Internet services have proved time and time again, the value of data increases with volume and diversity. The vision is to extend Internet technologies and experiences to sensors and thus create universal sensor-enhanced world connectivity and bring the Internet-scale service promise and capability to IoT/WoT.