IEEE IoT
IoT Scenario & Use Cases: Social Sensors

Service Description

More and more, people have the possibility to monitor important parameters in their home or in their surrounding environment. As an example, take the web site http://bwired.nl/, where each user can have a number of sensors monitoring and measuring parameters related to the functioning of the home or surrounding areas (e.g., the local outside temperature, the humidity, or even some parameters related to pollution, noise and others).

A Social Sensors service is intended to collect the wealth of user generated data, which can then be anonymized and compiled to show benefits for an entire community weighed against individual behavior. For example, it is possible to calculate a medium or average value for some parameters and allow each citizen to compare his own set of parameters with the "average set of values". In this way, individuals can gauge where they stand in respect to a set description of "virtuous citizen behavior", such as establishing a proper power consumption footprint. In fact, the availability of this type of data analysis could incentivize good-natured competition, and even encourage people to increase existing, or pursue new, practices perceived to be beneficial to the community as a whole.

Another possible usage is related to integrating user generated data in such a way to compare data and parameters collected directly by citizens versus official data provided by public administrations. One important case could be controlling local pollution vs. the official data monitored in a particular area of big cities.

Applicability Areas: Environment, E-Gov, Intelligent Home

The "Social Sensors" service aims to aggregate measures and information collected by sensors in a specific environment (e.g., a home) and to share them in a bigger context (e.g., a neighborhood). Data, measurements and information can be used for deriving knowledge (e.g., pattern analysis) related to how an environment is operating. An important feature is for users and owners of sensor networks to agree to share data in a larger community.. The service is characterized by the need and possibility to access and use data stemming from sensors in different administrative domains (e.g., homes, companies, public administration and government, social networks). This service sets an example of how large and independent sensor networks can cooperate in order to serve larger communities.

The Participants

- The **Sensors Provider** is the actor that actually manages a home sensor network that monitors many aspects of daily life. This includes things such as water and electricity consumption, to the frequency of people knocking at the door to deliver advertising, to many other activities centered on the "home". Data represents the behavior of the people living in a house, and storing and analysis of the data can give a very valuable description of how the social life of a family evolves over time.
- The **Aggregator** is the actor that actually collects and properly deals with a wealth of data. Its task is to try to govern the differences between the data representation and to try to organize data in a meaningful manner after it has been anonymized. The Aggregator could be also be a provider of a distributed sensor network, or the owner of other data collecting networks, such as utilities, network operators, or public administration/government agencies.
- The **User** ultimately takes advantage of all the information. The User could also be a company or programmer that is using the sourced data in order to determine the social behavior of citizens, or a very specific subset of individuals. (e.g., the inhabitants of a specific neighborhood).

A Social Sensors Service representation is provided in Fig. 1

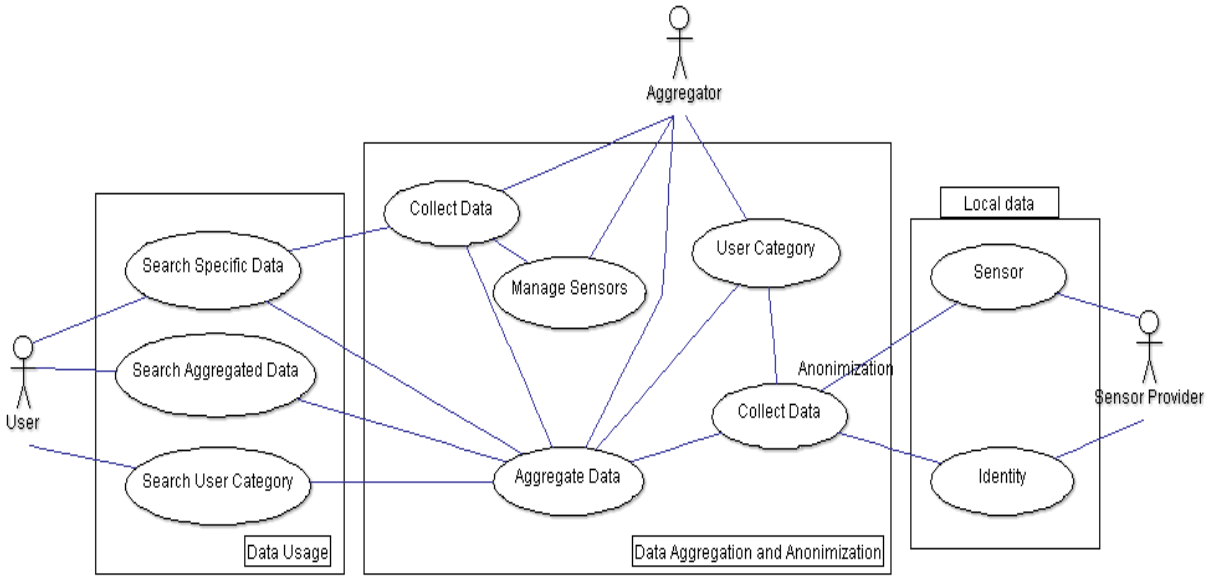

Fig. 1: A Social Sensors Service

The service is conceptually simple (see Figure 1): a Sensor Provider is a producer of information that is collected and anonymized by an Aggregator in order to harmonize the diversity in of the data format and information. Data collected from different Sensor Providers is aggregated and different views and/or inferred information can then be offered to a User. Data sets are also normalized in order to make them usable (i.e., different data formats can be generated and produced). The User can access the data or can utilize it as a

benchmark for behavior. The Aggregator could also integrate its own sensors infrastructure in order to create a wider set of data.

Applications range from very simple ones, such as using citizens' thermometer readings to determine the average city temperature, to very complex ones, such as using motion sensors or mobile operators' data to track a mob moving through a city.

While a social sensor service is conceptually simple, complexity is due to three factors. Firstly, there is a need to normalize heterogeneous sources (i.e., different sensors with different capabilities and different data representations and formats). Secondly, there is a need to anonymize data pertaining to the individual domain (i.e., data that allows a user to map a value of data to a specific user). Thirdly, there is a need to integrate data from and in different contexts and domains to address issues with communication, interworking and data reliability.

Social Sensors Business Model

A lot of different business models can be realized using the wealth of data generated by distributed sensor networks. It should be noted, however, that at least initially in the roll out of this service, opportunities and revenue models would not be totally clear. What's likely is that a Provider (e.g., a Utility) would deploy a wide area sensor network to measure power consumption homes, but that it could also integrate its sensor network in order to extend the basic service functionality by introducing such things as security and monitoring capabilities. The Aggregator could actually subsidize the deployment of wireless sensor networks in households, in return obtaining permission to track anonymized data and to use them for its business or for promoting new businesses.

It should be noted that it's extremely important to have a broad range of data generated by capillary distributed sensor networks. The competition between stakeholders will be to have expansive data sets from which to extract as much information as possible. This could result in better rules or agreements between Users and the Aggregator to assure User privacy. Awareness of what data are generated and passed to the aggregator can foster a user centric ecosystem in which the user is the owner of its profile. The monetization of user data is negotiated in a dynamic way according to policies stated and agreed upon by the user, and no longer imposed solely by the providers.

Figure 2 depicts a possible generic value chain in which different actors are identified.

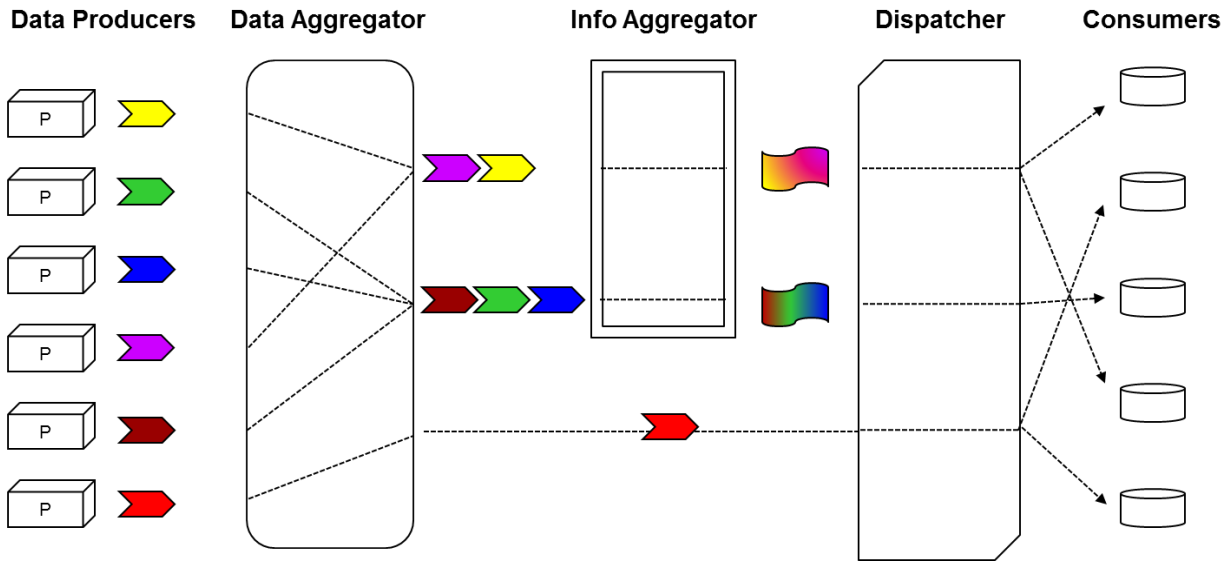

Figure 2: A Value Chain for the Social Sensors Service

The Data Producers (i.e.., the Users) make some raw data available. Data is aggregated and a part is directly dispatched (e.g., by means of PubSub engines) to consumers that will apply their data mining algorithms on the sets of data. Another possibility is for intermediate actors to extrapolate relevant information from data. Information is passed by a dispatcher (sometimes in real time) to consumers. The dispatcher is chargd with ensuring the delivery of the right information to the right consumers at the right time.

- Actors and their relationships
    - *Personal Consumer (e.g., elderly person in need of monitoring)* is a user that allows for the extraction/monitoring of some data from the data set produced by home sensor networks (Intelligent Homes).
    - *Providers of Intelligent Home Systems (Sensor Networks)* provide the technical solutions and the home infrastructure for monitoring some important parameters within the household (or the SME). The set of solutions available for the home will be mainly proprietary. There is the need to integrate data generated by different objects into a meaningful collection of comparable values.
    - *E-Gov Agencies* could benefit from the availability of data for promoting a better model of resources consumption in the home environment. They could support virtuous users/citizens by means of rewards or recognition of their good behavior (anonymity issue).
    - *Utilities* can favor the deployment of sensor networks in order to solve specific issues related to the working of their services (e.g., telemonitoring of gas/power consumption).

They can also support the increase in number of services and feature offered by the sensor network in the home (e.g., security features). This will allow them access to a wealth of data describing the anonymized behavior of a large number of users. They could use these data and others in order to characterize customers' behavior in order to better meet customer demand.

- *Service Provider*s are in charge of creating a platform able to collect, aggregate, normalize and anonymize data and make them (historically) available to applications. The Service Provider could develop a number of applications or create an enabling environment, such as a service platform with APIs.
- *Application developers* gain access to data and provide functions to the single customer (in a one to one relationship or contract) or to group of customers or even to a community of users.

IoT Architecture Requirements

This service stresses the data related capabilities of the IoT architecture. A first point is related to the way sensor data are collected, and assumes that a PubSub approach will be adopted. In this case, the dispatching of events and information has to be intelligently performed in order to satisfy consumers' requirements. Another issue is related to the storage and the organization of information. In this use case, it is assumed that data organization similar to that adopted by Twitter can be beneficial: a scenario where each single sensor can be virtualized in the infrastructure of the service provider. Each sensor will be individually identified and will generate a timeline of messages (very similar to tweets) that will be passed to the consumer and stored in the provider infrastructure. Different virtual sensors can be grouped to represent a set of sensors that operates in a specific room of the house (or even all the sensors in the house). The set of events from virtual objects' will generate a timeline that the owner of the sensor could control and use for its goals.

If sensors need to receive commands, a control message can be sent back to them, much like Twitter allows for the sending and receiving of private messages between different users.. Some of the virtual sensors could be anonymized, so that part of their timeline can be made available to aggregators who can use this information for running data mining algorithms to organize data or to extract information from the different timelines.. The aggregator has the need to access data in different "pages" (i.e., to check data in different sensors timelines) in order to dynamically extract the relevant information. This can be accomplished in a manner quite similar to Twitter, possibly using Hadoop or similar solutions. The coupling of PubSub and Hadoop extended with cognitive platform capabilities seems to be a strong indication for a workable IoT architecture. Figure 3 sketches a similar solution.

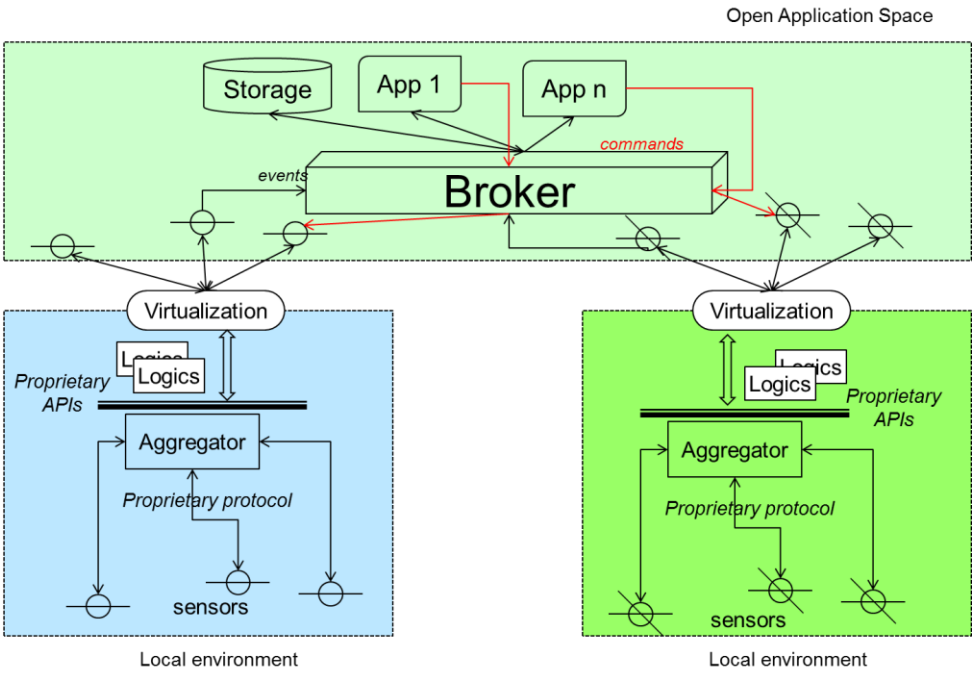

Figure3: A Possible Platform for Brokering of Social Sensor Derived Information

Putting together an efficient PubSub engine with a reliable mechanism to access distributed data (Hadoop) could enable an important role, that of a broker capable of efficiently and intelligently collecting, aggregating, inferring and dispatching information to consumers.

One important feature of the Social Sensors service is the ability to data mine large data sets. The pattern analysis seems to be relevant in this context because it could illustrate the behavior of people and groups of people when some important events occur (e.g., what is the change in behavior of people when an influenza epidemic is surging). This analysis can lead to models that describe how specific behaviors spread amongst the population.

Data management is clearly of the paramount importance in this example. The ability to aggregate data in a meaningful way can be a means for collecting interesting material for social studies. For instance, comparing the behavior of people in the same building, the same neighborhood, or the same city can lead to interesting insight as to how people tend to act in very similar ways and how they adapt and align to socially acceptable behaviors.

- Basic Functions supported by an IoT infrastructure
    - Data collection from different sources
    - Intelligent Event Dispatcher (e.g., PubSub based)
    - Integration of different formats and data representations
    - Storage and ability to retrieve historical data from a single source (a single sensor or a set of sensors)
    - Addition of sensors (i.e., a new resource) on a dynamic basis
    - Aggregation of sensors and monitoring functions for areas (e.g., in order to understand a phenomena in a certain area)
- Additional functions supported by an IoT Architecture

    - Anonymization of users associated with objects
    - Data Mining and Information Inference
    - Return of meaningful data to the single user and comparison of the single user with specific users (anonymous users or the “average” user)
    - Ability to allow access to historical usage data and creation of a sort of SPIME archive
- An indication of the functions/requirements that this use case is stressing
    - Representation of proprietary objects into observable and normalized virtual objects.
    - Anonymization of objects
    - Creation of a sort of social network for objects and ability to open up interfaces for data mining
    - Interoperability of objects pertaining to different administrative domains
    - Self-management with no user intervention