Heterogeneous Sensor Data Integration for Crowdsensing Applications

Sebastian Villarroya
CITIUS – University of Santiago de Compostela
Santiago de Compostela, Spain
sebastian.villarroya@usc.es

David Martínez Casas
CITIUS – University of Santiago de Compostela
Santiago de Compostela, Spain
david.martinez.casas@usc.es

Moisés Vilar
CITIUS – University of Santiago de Compostela
Santiago de Compostela, Spain
moises.vilar@usc.es

José R. Ríos Viqueira
CITIUS – University of Santiago de Compostela
Santiago de Compostela, Spain
jrr.viqueira@usc.es

José A. Taboada
CITIUS – University of Santiago de Compostela
Santiago de Compostela, Spain
joseangel.taboada@usc.es

José M. Cotos
CITIUS – University of Santiago de Compostela
Santiago de Compostela, Spain
manel.cotos@usc.es

ABSTRACT
This paper describes a conceptual solution for heterogeneous sensor data integration in crowdsensing applications and one experimental implementation for a health monitoring system in an educational environment using a low cost hardware solution. Three kinds of protocols are integrated in this solution: HL7 for medical data, Observations and Measurements model for environmental data and BACnet for buildings monitoring. This last protocol has the particularity that manages sensoring and acting. A Common Data Model is described for the integration of three kinds of data and protocols, and a validation test application is described.

Keywords
Heterogeneous data collection, Crowdsensing applications, Sensor data logging, Cloud Platform.

1. INTRODUCTION
Internet of Things is increasing exponentially its presence in our everyday-life and will have a high impact on potential users behavior, as exposed in [1], due to the integration of a five layer middleware architecture (applications, service composition, service management, object abstraction and objects) and identification, sensing and communication technologies. As stated in [2], many of the objects that surround us will be on the network. Such scenario provides a breeding ground for the deployment of mobile crowdsensing applications [3]. One of the most valuable domains we can find to apply the benefits of crowdsensing is healthcare. Near real-time monitoring of health state of individuals may provide valuable information in order to predict possible diseases and potentially contagious illnesses. Moreover, this information enables the study of spread patterns when diseases have already begun and is a useful tool when forecasting risk areas. Thus, using crowdsensing technologies we can allow individuals to provide valuable near real-time health information. The main objective of our work is to validate a heterogeneous sensor data-logger by developing an application that can collect data from the occupants of various educational centers to track common respiratory diseases among members of the educational community. This goal may be splitted into several specific issues such as collecting data of educational buildings using the BACnet[1] protocol, collecting environmental data of college areas from meteorological community services and facilities using the Observations and Measurements (O&M) data exchange protocol [4], collect health data from a low cost set of medical sensors using the Health Level 7 (HL7[2]) protocol, performing data analysis and providing a scalable solution.

The remainder of the paper is organized as follows. Related pieces of work are discussed in Section 2. The main objectives of this work and the common requirements in crowdsensing applications are exposed in Section 3. A Common Data Model (CDM) defined to integrate different data types involved in monitoring is presented in Section 4. The physical architecture of the platform and its implementation are described in Section 5. Finally, obtained conclusions and suggested future work are pointed out in Section 6.

2. STATE OF THE ART
The Cloud, mobile devices and sensors are the main components of the complex and heterogeneous scenarios within the context of Internet of Things. Such complex scenarios provide the opportunity for deploying crowdsensing applications within various domains such as healthcare and environmental data

IDEAS '14, July 07 - 09 2014, Porto, Portugal

http://dx.doi.org/10.1145/2628194.2628209

[1] http://www.bacnet.org

[2] http://www.hl7.org

monitoring. As an example of this effort, in [5] an air quality monitoring system is developed as a use case for a publish/subscribe middleware defined in the field of energy-efficient mobile crowdsensing. A schema to convert data between HL7 devices and monitoring devices in a wireless home healthcare scenario is proposed by [6]. Similarly, our platform provides data integration by adapting different data types to the CDM (our common data model) observation types. This adaptation is possible using different data acquisition channel wrappers. An abstract specification of a data model for Observations and Measurements (O&M) [4] in a geographic information context is defined by the Open Geospatial Consortium (OGC). Data acquisition systems for monitoring purposes, which is other important functionality goal of our platform, are analyzed next. A framework for the development of data acquisition and dissemination servers called GeoDADIS is presented in [7]. Data acquisition and dissemination are main objectives of our platform as well. Another similarity is the adoption of the Adapter pattern [8] to access the specific interfaces of data acquisition channels and data services in a uniform manner. It provides the flexibility to acquire data and serve information using any communication protocol by adding specific wrappers. Unlike GeoDADIS, our platform does not provide in situ data acquisition. Spatio-temporal stamping of sampled observations is out of the scope of our platform. A common data model, that is not present in GeoDADIS, is defined in our platform to provide a homogeneous representation of heterogeneous datasets.

3. OBJECTIVES OF OUR WORK

The main objective of this work is to develop an application that can collect data from the occupants of various educational centers to track common respiratory diseases among members of the educational community. Among other methods, we use crowdsensing (a capability that harnesses the power of crowds to collect sensor data from a large number of users) to collect data.

To achieve the main objective we have a number of secondary objectives:

– Collect data of educational buildings through the BMS (Building Management System) using the BACnet protocol. These dataset are composed by measures of environmental parameters of the buildings such as indoor temperature, indoor relative humidity and building occupancy.

– Collect environmental data from the colleges surrounding areas obtained from some of the meteorological community services using the O&M data exchange protocol. These data are external environmental measures such as outdoor temperature, relative humidity, air quality, wind speed and direction.

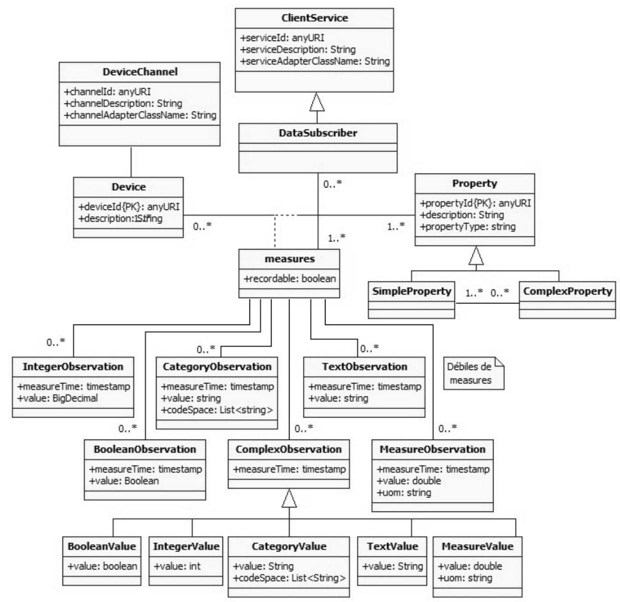
– Collect health data, through the HL7 protocol, using a low cost set of medical sensors available in schools: nasal airflow sensor to determine breathing status, body temperature sensor to determine fever presence, pulse and blood oxygen sensor.

With all the collected data we can do several analysis such as correlation studies, to determine which are the best parameter configuration of the buildings to minimize the risk of respiratory diseases and maintaining good health status of the building occupants.

BACnet is also used to set the buildings parameters as our system is able to manage both sensors and actuators.

4. COMMON DATA MODEL

Our application requires a common data model that allows us the integration of different types of data input arriving at the platform: environmental, medical or industrial. To achieve this we develop a data model to unify data from very different data sources: BACnet, O&M and HL7.



ClientService. Communication channel between the platform and the applications fed by platform data, both in synchronous or asynchronous manner. This provides, through a set of wrappers, the flexibility to cover a large set of technologies.

DeviceChannel. This element of the model represents the communication channel to communicate devices with the system. This element takes account of the various data communication protocols used by the sensors that send data to the system such as TCP, HTTP and MQTT.

Device. A device element of the model represents a mechanism that can perform measurements on physical magnitudes, i.e. a sensor.

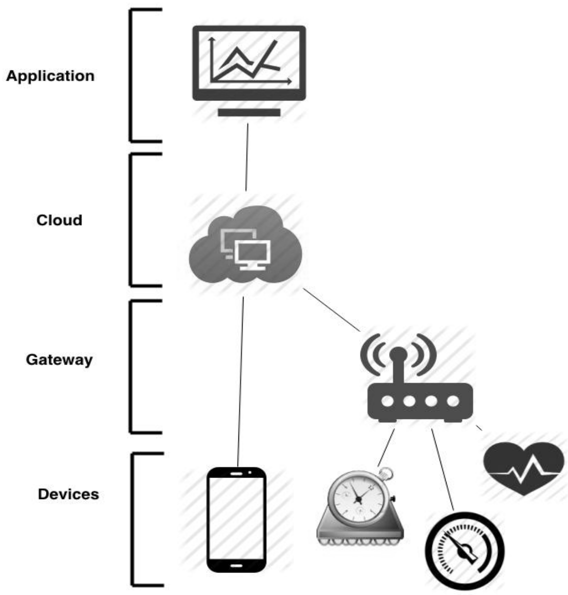
Property. A property represents a measure attribute that can be measured by a device and referenced by a name. There are two property categories: SimpleProperty, property represented by simple observations, ComplexProperty, set of simple properties.

Measure. Association between device and property to take into account that a device can measure various properties and a property can be measured by various devices.

Observation. Represents the value of a property obtained through a measurement process. An observation has a reference to the device that has performed the mea-surement process, a reference to the measured property, the time instant of the measurement process and the value of the measurement result. Depends on the type of result there are several observation types: BooleanObservation (true/false), CategoryObservation (option in a set), IntegerObservation (dimensionless integer), MeasureObservation (real number with unit of measure), TextObservation (free text) and ComplexObservation (aggregation of any of the above).

5. ARCHITECTURE

The physical architecture of the platform we have developed is organized around a schema in four levels



5.1 Devices

To capture the data we use various sensor devices to monitoring the buildings (BACnet), the environment (O&M) and the user's medical data (HL7). To develop our application we have use a set of low cost sensors:

Smartphone sensors. We can use the sensors usually have modern smartphones to gather information of interest. With accelerometers we can infer the type of activity of the user (sitting, walking, running, ...) and with the GPS the user's position.

Health sensors. We can collect user's health status through low cost specific sensors like airflow sensor, body temperature sensor, pulse and blood oxygen sensor and manual status data.

Environmental sensors.

– Buildings. BACnet allows us to collect parameters of buildings such as indoor temperature, humidity or occupancy.

– Environment. Through state agencies[3] we can obtain environmental data near the educational buildings under study; data such as weather conditions or outdoor air quality.

5.2 Gateway

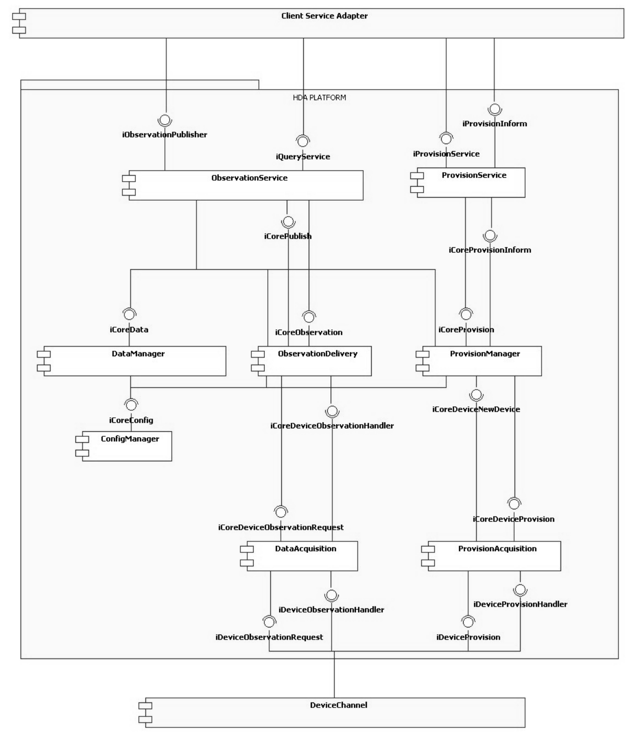
We have developed a multi-platform architecture that is hardware independent, and made a deployment in a Raspberry Pi[4]; a credit-card-sized single-board low cost computer.

The gateway is used to establish the communication with the sensors, the communication can be synchronous or asynchronous. When sensor data reaches the gateway the data is stored in a tiny local database H2[5] and send to subscribers through a publish/subscribe mechanism.

The gateway also has enough computing power and is capable of certain pre-processing or validations on the data.

5.3 Cloud

A cluster infrastructure based on a 6 PC cluster was deployed to collect the data of all the gateways. The cloud has the UML module schema shown in the figure.



DeviceChannel adapter. This interface specifies how device adapter connects to the platform to send data, both in a synchronous or asynchronous mode.

ClientService adapter. This interface specifies how applications can consume data from the platform, both synchronous and asynchronous.

IObservationPublisher. This interface manages the notifications to the services that are subscribed to the concrete observations through publish/subscribe paradigm.

IQueryService. This interface manages the queries that applications perform over the observations stored in the cloud database.

IProvisionService. This interface is responsible to manage metadata to provide a device provision service. It provides Create, Read, Update and Delete (CRUD) operations over data model entities.

IDeviceObservationHandler. This interface is responsible for entering sensor data into the platform from gateways.

[3] http://www.meteogalicia.es/

[4] http://www.raspberrypi.org/

[5] http://www.h2database.com/

5.4 Application

The main objective of the application is to test the platform described in this paper. This application is fed by the platform data and performs a set of analysis to reduce the chances of transmission of respiratory diseases in secondary schools. In the next section the prototype implementation will be described more extensively.

The purpose of the application is divided into the following points:

- Collect the data from the devices through the platform.
- Perform a particular analysis on the data.
- With the results of the analysis give a set of recommendations and change the building parameterization to minimize the chance of infection on respiratory diseases.

Using the above functionality we provide a set of benefits to all of the data contributors:

- Recommendations to improve user's health and to minimize the chance of contagious.
- State determination and activity quantification of users, buildings and environment.

With all the heterogeneous data captured through the platform a series of analysis can be performed:

- Spread analysis of respiratory contagious diseases: influenza.
- Correlation study between illness and building and environment conditions.
- Activity analysis: correlation between user's activity and respiratory disease.

Nowadays, a real crowdsensing application has not being deployed yet. We only have tested a prototype at laboratory level.

As we do not have any device to provide information about whether or not a user suffers from flu, we will look primarily on two parameters to determine its presence:

- The user body temperature, to determine the presence of fever.
- The user status described manually by himself through a form provided by the application.

In order to have discretized data samples, data of users for which we do not have some certainty about whether or not are suffering from flu will be removed from the database.

The next step should be to analyze the correlation between the presence or non-presence of influenza in the users with the values of all other parameters to which we have access over the platform. Doing this we would try to isolate the possible causes that have greater influence on the occurrence of the flu disease. After experts consultation on respiratory diseases and correlation analysis result, a report would be created with the most appropriate actions to minimize the incidence of the disease.

6. CONCLUSION AND FUTURE WORK

Through the work described in this paper we have reached several conclusions:

The platform that we developed has the ability to collect data from various sources: health, indoor environment, outdoor environment, ... This data can be acquired by the platform using various data protocols: BACnet, O&M and HL7.

Once we have all the data collected under a unified data model (CDM) the process of analyzing them is pretty easier. In addition the unified model permit us give back data to all those involved in the crowdsensing process to do their own studies.

The results of the analysis provided by the application can be delivered in a report to the educational managers and them they can take actions such as changing the building parameterization to reduce the spread rate of the disease.

The most interesting issue in the near future is the implementation of a system of knowledge-based rules to alert us of certain situations of interest, such as the values of certain parameters exceed a defined threshold or a combination of several parameters meet certain condition. The main objective is to be able to extract information to do further data analysis and act quickly on the configuration of the buildings to improve occupant's health.

Another interesting point is the development of virtual processes, i.e. measurements obtained by combining two or more physical measurements. These virtual processes can be used as virtual measures with more condensed information or as custom formatted measures for any specific purpose.

7. REFERENCES

[1] L. Atzori, A. Iera, and G. Morabito. The internet of things: A survey. Computer Networks, 54(15):2787 – 2805, 2010.

[2] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami. Internet of things (iot): A vision, architectural elements, and future directions. Future Generation Computer Systems, 29(7):1645 – 1660, 2013.

[3] M.-R. Ra, B. Liu, T. F. La Porta, and R. Govindan. Medusa: A programming framework for crowd-sensing applications. In Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services, MobiSys '12, pages 337–350, New York, NY, USA, 2012. ACM.

[4] Open Geospatial Consortium (OGC). http://www.opengeospatial.org/standards/om.

[5] I. Podnar Zarko, A. Antonic, and K. Pripuˇ zic. Publish/subscribe middleware for energy-efficient mobile crowdsensing. In Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication, UbiComp '13 Adjunct, pages 1099–1110, New York, NY, USA, 2013. ACM.

[6] M. Lee and T. M. Gatton. Wireless health data exchange for home healthcare monitoring systems. Sensors, 10(4):3243–3260, 2010.

[7] S. Villarroya, J. Viqueira, J. Cotos, and J. Flores. Geodadis: A framework for the development of geographic data acquisition and dissemination servers. Computers & Geosciences, 52(0):68 – 76, 2013.

[8] E. Gamma, R. Helm, R. Johnson, and J. Vlissides. Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley, NJ, USA, 416pp, 1995.