Big Data Platform for Monitoring Indoor Working Conditions and Outdoor Environment

Conference Paper · November 2014

Igor Mishkovski
Ss. Cyril and Methodius University in Skopje

Lasko Basnarkov
Ss. Cyril and Methodius University in Skopje

Ljupco Kocarev
Macedonian Academy of Sciences and Arts

Svetozar Ilchev
Bulgarian Academy of Sciences

Big Data Platform for Monitoring Indoor Working Conditions and Outdoor Environment

Igor Mishkovski
Faculty of Computer Sciences and Engineering, P.O. Box 393, 1000 Skopje, R. Macedonia.
e-mail: igor.miskovski@finki.ukim.mk
Lasko Basnarkov
Faculty of Computer Sciences and Engineering, P.O. Box 393, 1000 Skopje, R. Macedonia.
e-mail: lasko.basnarkov@finki.ukim.mk
Ljupcho Kocarev
Macedonian Academy of Sciences and Arts. Bul. Krste Misirkov 2, P.O. Box 428.
e-mail: lkocarev@manu.edu.mk
Svetozar Ilchev
Institute of Information and Communication Technologies – BAS, Acad. G. Bonchev Str., Bl. 2, 1113 Sofia, Bulgaria
e-mail: svetozar@ilchev.net
Rumen Andreev
Institute of Information and Communication Technologies – BAS, Acad. G. Bonchev Str., Bl. 2, 1113 Sofia, Bulgaria
e-mail: rumen@isdip.bas.bg

Abstract: This paper presents a framework and a test bed for monitoring indoor working conditions and outdoor environment. In particular, we focus on developing: i) a test bed small Wireless Sensor Network (WSN) for monitoring indoor and outdoor environment parameters using low cost Arduino controllers, transceivers and sensors, and ii) real-time platform for analysis and mining of the sensor data. Together the two integral parts will offer not only monitoring, but also mining of the data and detection of any environmental anomalies.

Keywords: Big Data, Sensor, WSN, Mining

1. Introduction

The environment changes quickly, and these changes influence the citizens' health, perceived quality of life and work efficiency. Environmental changes also influence the economy directly e.g. tourism and agriculture. Thus, a great part of the research community is still searching for the right kind and amount of data and analysis tools necessary to address serious problems that occur unexpectedly and develop rapidly. Furrthermore, the devices with sensing capabilities are becoming ubiquitous, e.g. low-power sensor networks or mobile and wearable devices equipped with sensors. On the other hand, data mining, machine learning are now able to deal with large-scale data sets that contain millions of high-dimensional data points.

The proper monitoring of the outdoor environment and alerting when certain anomalies arise, addresses the major environmental health treat [1]-[5]. However, outdoor environment monitoring, will not only influence the public health, but also, the quality of life and the working efficiency of the citizens.

On the other hand, monitoring the indoor working conditions can improve health, work performance and school performance, reduce health care costs and be a source of substantial economic benefit [6]-[11].

The collected data, both outdoor and indoor, can be collected in data storage and used together with the real-time stream of data for visualizing, anomaly detection and building a model for prediction of employee productivity, etc. In order to process all the sensor information, aggregate it and disseminate it back to the users in relevant way a central Real-Data Processing System will be required, as in [12]. This system will queue and map heterogeneous data and will serve as a service real-time layer for different Distributed Remote Procedure Calls (DRPC). Besides offering different open web services, the platform can offer semantically annotated linked data.

The test bed Wireless Sensor Network (WSN) will statically monitor the environment parameters, as well as it will address mobile monitoring of the condition in the outdoor environment. The collected data will be stored on the Real-Data Processing System, from which we will do indoor and outdoor layered visualization on the building plan and/or GIS systems, respectively. Using the stored data and the real-time data the system will infer anomalies, as well as, using a measure of the employee productivity it will learn which are the conditions that increase the employees' productivity.

Thus, in this work we propose a simple, open and cheap framework that offers data services, both raw and processed. This data allows modelling and exploration of the relations between variables in environment and detection of alarming trends. The platform also will provide feedback to businesses and citizens about influence of their actions on the environment.

The paper is organized as follows. In Section 2 we give the overview of the proposed sensing framework and the possible sensors that could be used for monitoring indoor working conditions and outdoor environment. Section 3 gives overview of the Real-Data Processing System and Section 4 concludes this paper.

2. Architecture for Indoor and Outdoor Monitoring

The diagram in Figure 1 shows the possible architecture for indoor and outdoor monitoring. The architecture is consisted of several Arduino static nodes that will monitor the indoor conditions. In the diagram, the architecture will measure several parameters, such as Pressure (P), Temperature (T), Humidity (H), Dust and data from other possible sensors. The sensor nodes will continuously read the status of all attached sensors and pass the sensor data through the radio network back to the gateway. These sensors will have the option to sleep most of the time in order to save battery. However, in the system there might exist repeater-sensor nodes (not shown in Figure 1) which must stay awake in order to pass messages from their child sensor nodes. A repeater-node can optionally include direct-attached sensors and report their sensor data to the gateway.

The Arduino sensor nodes will communicate with the Arduino Gateway using the NRF24L01+ transceiver from Nordic Semiconductors which communicates with the Arduino board via the SPI interface. The Arduino Gateway on the other hand will act as a glue between the controller and the radio network. It will translate radio messages to a protocol which can be understood by a controller. There are several possible implementations for the gateway:

- SerialGateway - The gateway connects directly to the controller using one of the available USB ports.
- EthernetGateway - The gateway connects to the Ethernet network that the controller also uses offering more placement flexibility than the SerialGateway.
- MQTTGateway - This gateway also connects to the Ethernet network and exposes an MQTT broker which can be used for controllers offering MQTT support like OpenHAB [13].

AS a good candidate for the controller we will develop our own simple DIY cloud-enabled gateway controller running on the Raspberry Pi.

Besides the serial communication between the controller and the Arduino gateway the controller will collect FTP Data sent by the GPRS module of the mobile Arduino sensor nodes. The mobile Arduino sensor nodes, besides the GPRS module will be equipped with additional sensors for P, T, H, Shinyei PPD42NS Particle sensor, GPS sensor, and some other possible sensors. All the sensor nodes will be boxed in special cases using 3D printer.

The controller will collect all the indoor and outdoor sensor data through the serial and ftp communication, respectively and will feed the data into the real-data processing system.

2.1. Other Relevant Sensor Data and Sensing the Employees' Productivity

Beside the abovementioned sensors, the system can be upgraded with additional sensors that can fed more data into the real data processing system, such as:

- Open/closed doors or the state of a wall switch.
- Distance sensor - it can measure the sitting habits of an employee.
- Gas sensor - for detecting alcohol, methane, fire, etc.
- Infrared sensors - that can control the air-conditioning system in the office.
- Light sensor - can be used in the automated control of the drapers.
- Movement sensor - to detect if the office is overcrowded and what are the dynamics in the office.
- Relay Actuator - to turn on/off the devices.
- RFID sensors – to detect the workers in the office.
- Infrared sensor.
- Noise meter – to measure the noise conditions in the office.
- UV sensor – to measure the UV factor in the environment.

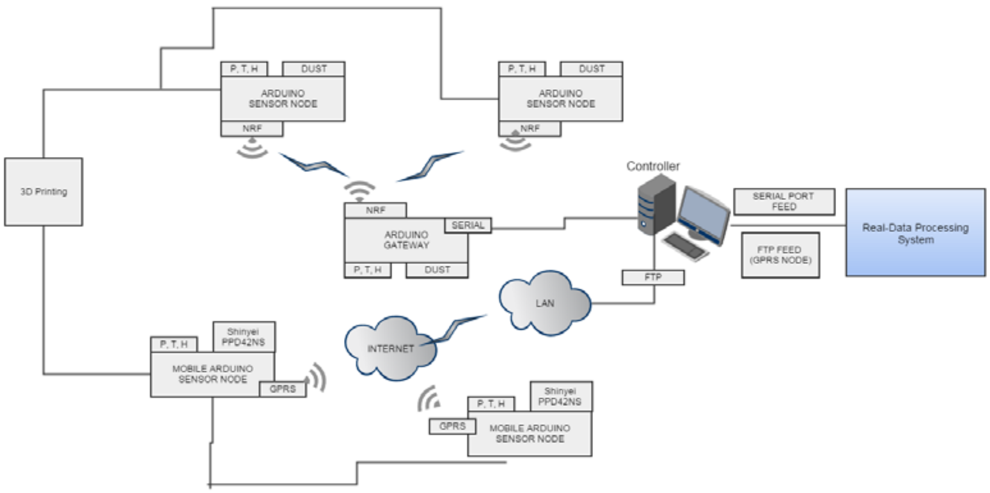

Figure 5 Architecture for Indoor and Outdoor Monitoring

In order to measure the employee productivity we propose four distinct way that will be used as an output feature of the machine learning techniques, such as:

- Using user input – with the help of a button input (productive/non-productive day)
- Measure productivity by gathering the production data, i.e. how many pieces or products were produced in one day in some factory
- Sensor on chairs – using this sensing data we will obtain information how much and what are the sitting habits of the employee.
- Use the data from some corporate task platform (such as google task, or some proprietary).

3. Real Data Processing System

The real-data processing system will be fed by the controller and it provides an architectural model that scales and which has both the advantages of long-term batch processing and the freshness of a real-time system, with data updated in seconds' time.

The data will be are fed into the system (1), for example through a queue from where the system (such as Storm) can pull them. A system, such as Trident, will save them into Hadoop (HDFS) and processes them in real-time for creating an in-memory state. In Hadoop all the historical data will be available, and in any time a batch process can be started that will aggregate the data and generate a big file from it. After that, we can use some API tools or SQL command line tools (such as Splout) to index the file and deploy it to a Splout SQL cluster (4), which will be able to serve all the statistics pretty fast. Then, a second stream (DRPC), such as Trident, can be used to serve timeline queries, and this stream will query both the batch layer (through Splout SQL) and the real-time layer (through the first stream's memory state), and mix the results into a single timeline response. In this way, we will prepare both the historical and the real data for the stakeholders in order to visualize it, receive alerts, and do some possible real-time optimizations.

4. Conclusions

The proposed platform, based on data collected from various sources and processed by online services, should help decision makers in finding balance, optimal trade-offs in real-time. Moreover, using this simple and cheap platform, the stakeholders not only that can obtain various information about the indoor and the outdoor conditions, they can create a model for the employee productivity depending on the conditions and certain anomalies in environment that affect the production. Finally, the employers can affect some of the indoor conditions in order to boost the employee productivity using the obtained models.

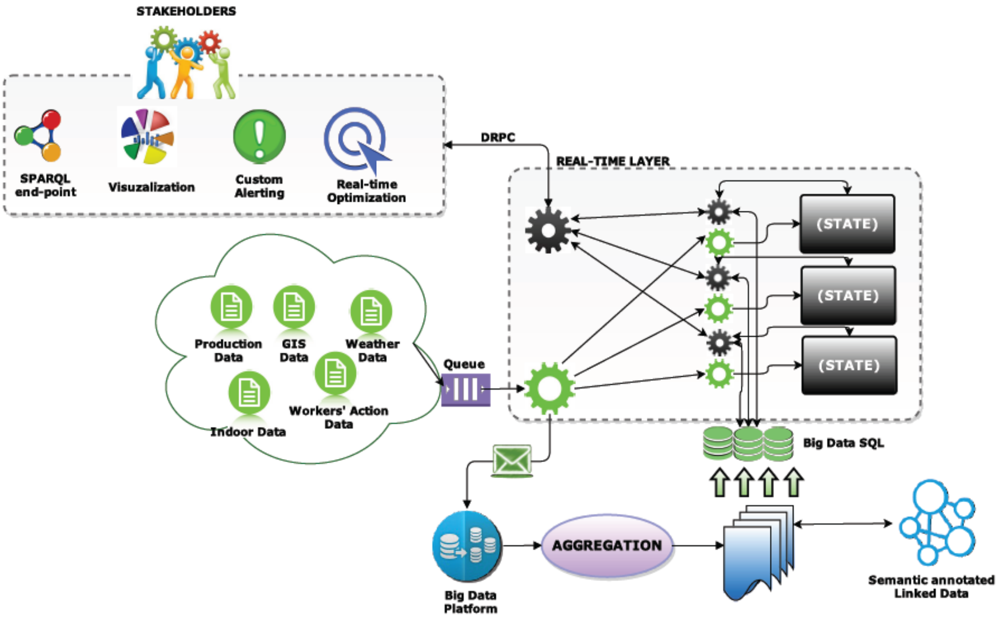

Figure 6 Real-data processing system

References

[1] Kampa, M. & Castanas, E. 2008. Human health effects of air pollution, Environmental Pollution 151:362-367.

[2] Brunekreef, A. & Holgate, S.T. 2002. Air pollution and health. Lancet 360:1233-1242.

[3] Dockerty, D.W., Arden Pope, C., Xu, X., Spengler, J.D., Ware, J.H., Fay, M.E., Ferris, B.G. & Speizer, F.E. 1993. An Association Between Air Pollution And Mortality In Six U.S. Cities. The New England Journal of Medicine 329(24):1753- 1759.

[4] Pope C. A. III, Burnett, R.T., Thun, M. J., Calle, E.E., Krewski, D., Ito, K. & Thurston, G. D. 2002. Lung Cancer, Cardiopulmonary Mortality, and Long-term Exposure to Fine Particulate Air Pollution. Journal of the American Medical Association 287:1132-1141.

[5] Mucke, H.-G. 2000. Ambient air quality programmes for health impact assessment in the WHO European region, Arh Hig Rada Toksikol 51:257-564.

[6] ASHRAE(2011) ASHRAE Position Document on Indoor Air Quality, Atlanta, GA, USA, American Society of Heating, Refrigeration, and Air-Conditioning Engineers, Inc.

[7] Schell, M. and Inthout D. 2001. Demand Controlled Ventilation Using CO2, ASHRAE Journal.

[8] Stranger, M., Potgieter-Vermaak, S.S., Van Grieken, R. 2008. Characterization of indoor air quality in primary schools in Antwerp, Belgium. Indoor Air 2008; 18:454- 463.

[9] Currie, J., Hanushek, E.A., Kahn, E.M., Neidell, M., and Rivkin, G. 2009. Does Pollution Increase School Absences? The Review of Economics and Statistics, November 2009, 91(4): 682-694.

[10] http://www.cdc.gov/HealthyYouth/asthma/

[11] Maclean, M.; Anderson, J.G.; MacGregor, S.J.; Mackersie, J.W. 2004. "The development of a pulsed UV-light air disinfection system and its application in university lecture theatres," Power Modulator Symposium, 2004 and 2004 High-Voltage Workshop. Conference Record of the Twenty-Sixth International , vol., no., pp. 630- 633, 23-26 May 2004.

[12] Rizea, Daniel-Octavian, Olteanu, Alexandru-Corneliu, Tudose, Dan-Stefan. 2014. "Air quality data collection and processing platform", RoEduNet Conference 13th Edition: Networking in Education and Research Joint Event RENAM 8th Conference, Moldova, September 2014.

[13] openHAB, http://www.openhab.org/.