Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 113 (2017) 429–434

Procedia Computer Science

www.elsevier.com/locate/procedia

The 7th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH 2017)

On the use of IoT and Big Data Technologies for Real-time Monitoring and Data Processing

Y. Nait Malek[a,b], A. Kharbouch[b,c], H. El Khoukhi[b,c], M. Bakhouya[a,*], V. De Florio[d,c], D. El Ouadghiri[b], S. Latre[c], C. Blondia[c]

[a]*International University of Rabat, Faculty of Computer Science and Logistic, Technopolis, Sala Aljadida, Morocco*

[b]*Université My Ismail, Faculté des sciences 11201 Zitoune, Meknès - Morocco*

[c]*Universiteit Antwerpen (UA) and Imec, Mathematics and Computer Science Dept. Middelheimlaan 1, 2020 Antwerp, Belgium*

[d]*Vlaamse Instelling voor Technologisch Onderzoek, Belgium*

Abstract

Recent advances in pervasive technologies, such as wireless ad hoc networks and wearable sensor devices, allow the connection of everyday things to the Internet, commonly denoted as Internet of Things (IoT). IoT is seen as an enabler to the development of intelligent and context-aware services and applications. These services could dynamically react to the environment changes and users' preferences. The main aim is to make users' life more comfortable according to their locations, current requirements, and ongoing activities. However, handling dynamic and frequent context changes is a difficult task without a real-time event/data acquisition and processing platform. Big data and IoT technologies have been recently proposed for timely analysing information (i.e., data, events) streams. In this paper, we propose to combine IoT techniques with Big data technologies into a holistic platform for continuous and real-time data monitoring and processing. Preliminary experiments have been conducted and results are reported to show the usefulness of this platform in a real-case scenario.

Peer-review under responsibility of the Conference Program Chairs.

Keywords: Internet of Things, Big data technologies, Complex event processing, Real-time monitoring and processing, Context-awarness.

[*] Mohamed Bakhouya. Tel.: +212-053-010-3070; fax: +212-053-010-3100.
E-mail address: mohamed.bakhouya@uir.ac.ma

1877-0509
Peer-review under responsibility of the Conference Program Chairs.
10.1016/j.procs.2017.08.281

1. Introduction

Advanced mobile and wireless ad hoc networks and the large deployment of things (e.g., sensors, actuators, agents) in buildings and homes, in public facilities, and in outdoor environments allow the development of new context-aware applications and services. In other words, the connection of everyday things to the Internet commonly denoted as Internet of Things (IoT) is often seen as an enabler for developing new intelligent applications and services. The IoT could be built on the pervasive deployment of a variety of things, such as RFID tags, sensors, actuators, mobile phones that are able to interact with each other and cooperate with other services to reach common goals[6]. In fact, the integration of data generated by this myriad of sensors with other data, such as location, environmental context and social media data allows the development of context-aware applications that could help citizens taking more informed decisions regarding their day's activities. For instance, it is now possible to develop services for better traffic routing throughout the city, detect and immediately act to environmental pollution peaks or automatically optimize the logistics chain by allowing instantaneous reactions to external triggers and contextual changes.

However, the growth of these streaming data creates what is named the "Big Data" phenomenon with the main four V's, Volume, Velocity, Variety and Value[2]. Recently, several IoT and Big data technologies have been developed to allow the collection, the processing and analysis of large amount of streaming data. The aim is to timely extract the ongoing and future contexts in order to anticipate the changes and tailor services that better fit the users' needs. Several IoT platforms have been proposed for easy deployment of context-aware applications. However, current IoT applications, such as Thingspeak[23], can be used for data monitoring (e.g., traffic monitoring) and analysis without any guarantees on transmission and processing. A consequence of this is that streaming the data from an embedded device into a data analytics cloud and performing analysis based on this data can easily take tens of seconds.

In parallel to these progress in IoT platforms, several Big Data technologies have been developed and could be classified according to the data processing concepts used[3,4]: *batch processing technologies* and *real-time processing technologies*. Batch processing technologies, such as Hadoop[5], are more suitable for high throughput data processing. The principle used is to first store data and process it later on. However, real-time processing technologies, such as Storm and S4, have been developed to process data in motion and get as fast as possible valuable insights from it. The aim is to enable the development of applications that require real-time or near real-time processing. These applications should operate in scenarios with tight requirements, such as high data rates (more than 10000 event/s) and low latency (up to a few seconds) [7].

In this paper, we introduce the architecture of a holistic platform by combing tools from IoT, complex event processing and Big data technologies. The aim is to provide a platform that allows processing large scale sensor data in order to develop context-aware applications and services. We mainly based our study on the pioneering work aiming at combining KAA and Storm technologies into a holistic platform[9]. In fact, Storm is a distributed event stream processing[8] that allows distributing the processing and input data streams into several processing units in order to reach near real-time and make mitigation action as quick as possible. While Storm supports distributed real-time computations, Kaa in turn is an open-source middleware tool for building, managing and integrating various IoT devices. A prototype of the platform has been developed to show its usefulness using a small healthcare scenario. The remainder of this paper is structured as follows. Section 2 presents work dedicated to real time processing using recent advanced IoT and Big data technologies. In Section 3 an overview of the platform architecture is presented. The case study scenario is described in Section 4. Conclusions and ongoing work is given in Section 5.

2. Related work

Recent advances in sensing technologies, such as embedded and wearable devices, and their interconnection with wireless networks has led to the deployment of context-aware wireless sensors networks. The integration of these wireless networks with Internet allows the emergence of IoT[19] infrastructures that produces high volume of streaming data. The availability of those data makes development and deployment of context-aware applications and services possible, such as those in smart grids, smart homes, and transportations. These applications, for instance, will allow users to receive real-time feedbacks about their context, enabling them to make their own informed decisions and mitigation actions. However, these advances create large challenges related to data monitoring and processing.

Several initiatives have been recently introduced by developing real-time processing platforms and approaches that could handle the requirements of these applications. The two major communities are: *complex event processing* and *distributed data stream processing*. Researchers from the first community are developing technologies for high throughput input/output data processing[3,4]. There are mainly two processing technologies, *batch processing technologies* and *real-time stream processing technologies*. Batch processing technologies are based on the first store and process principle, which has latency measured in minutes or more like in MapReduce[11]. However, real-time stream processing technologies, such as Storm, have been developed to process data while in motion in order to get as fast as possible valuable insights from it.

In parallel to this progress in Big data technologies, researchers from complex event processing community have been developing expressive languages and techniques for advanced pattern matching (e.g., complex events) [12,13,14]. Examples of recent and extensively used engines are C-SPARQL, ETALIS, and CQELS[15,16,17]. However, most of the works to date have been focused on evaluating, using benchmarks, the scalability of these engines in handling high events streams rates. For instance, recent studies showed that ETALIS outperforms other engines especially for scalability[14,18]. However, their deployment for real-time processing is still difficult without their efficient integration with Big data technologies. More precisely, main context-aware applications require scalable platforms that could handle large amount of incoming data streams as well as techniques for extracting relevant patterns according to changing situations. Therefore, a platform combining features from complex event processing and streaming processing from big data field is required for extracting and reasoning about systems and environments context[14].

Recent research directions put more emphasis on the development of integrated platforms for real-time data monitoring and processing. For instance, authors[10] have designed and implemented the VISP ecosystem for data stream processing for IoT. It is composed of a Marketplace component that allows users creating IoT-based topologies and a runtime component for the execution of these topologies. Our work put more emphasis on the integration of recent state-of-the-art IoT and Big technologies by introducing an integrated platform using KAA[10] and Storm for collecting, processing, and analyzing large scale sensors data. A prototype was developed and preliminary results are reported to show its usefulness for developing context-aware services.

3. Platform architecture

This section presents the architecture of the platform developed for processing real time streaming data from sensors. We choose to use open source tools like Apache Storm, Kaa and Apache Flume[25]. The platform is composed of three layers as follows.

- **Data acquisition**: Kaa platform is used for gathering data from various sensors. This platform is an open-source middleware for building, managing and integrating various IoT devices. It is horizontally scalable, fault tolerant and provides a several IoT features.
- **Data processing:** among all existing processing systems, we choose apache Storm, because of its ability to process real time data streams. It is free, open source, simple and can be used with several programing languages.
- **Data storage and visualization:** MongoDB is used for storage and a web application is developed for data visualization. Processed data could be then stored for eventual use (e.g., long terms data mining) or for visualization.

The aim of this new platform design is to integrate new dedicated HW/SW that will mainly be conceived to shift a part of data processing into the in-field devices. This will enable the scalability of the platform by lowering remote processing time, with low data transmission and storage. We are targeting a more efficient hardware/software approach in order to improve real-time constraints in many context-aware applications and services. These HW/SW-oriented devices will be efficiently interfaced with IoT technologies (e.g., KAA

) and Big Data platform (e.g., Storm, S4) for real-time data processing. Fig. 1 depicts a schematic view of the prototype components we have deployed and tested for usefulness in developing context-aware services. It is composed of main three components, *sensing nodes*, *in-field processing devices*, and *processing/storage infrastructure*.

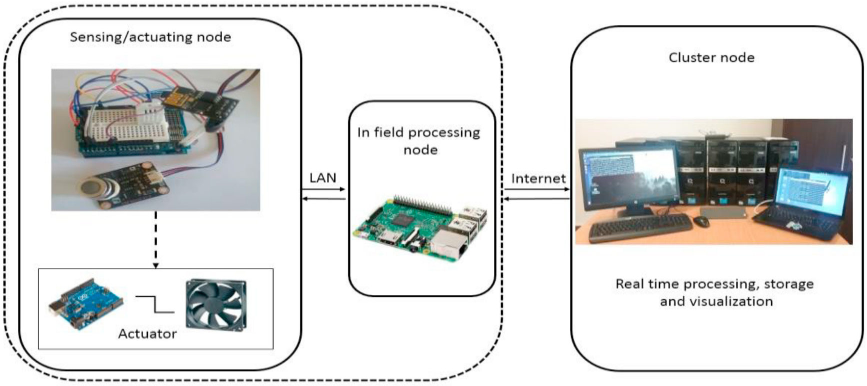


Fig. 1. The prototype platform architecture

Sensing nodes are Arduino micro-controllers that embed related sensors and actuators. Sensors are used to gather in-field monitored values and send them, using wireless communication modules (e.g., ESP8266), into the in-field processing device. For communication, we used MQTT[24] (Message Queue Telemetry Transport) to communicate data between the Arduino with the Kaa applications. MQTT is a machine-to-machine connectivity protocol. It is designed as an extremely lightweight publish/subscribe messaging transport. In our first platform prototype, we have used a Raspberry devise for short terms processing (e.g., aggregating data). It incorporates the Kaa client application that gets sensors data from the Arduino program. These processed data are then submitted to the main node of the Storm cluster, which embeds the application according to the processing topology. The topology is built using the spout/bolt entities[8,9]. Spouts represent sources of streams, and bolts represent the operators for data correlation and computation. In order to transmit data from Kaa applications to Storm applications, Apache Flume is used. It is a distributed and reliable service for collecting, aggregating and moving large amounts of data; it has a simple and flexible architecture based on streaming data flows. The cluster-based infrastructure is used for long data processing and storage. In this first prototype (see Figure 1), we have used a cluster composed of six nodes.

4. Real-case scenario and experiments

The purpose of this real-case study is to examine and analyse vital signs of buildings' occupants when exposed to pure carbon dioxide and bioeffluent. We have used non-invasive biomedical sensors for gathering vital data without penetrating the body. In our experiments, we used open hardware sensors, such as pulse and oximetry, carbon dioxide in air, humidity and temperature sensors. The aim is to study how the lack of proper building's ventilation can impair occupants' performance and affect their health. We have used indoor CO_2 concentration as an indicator of air quality. It was used as an indicator for controlling buildings' ventilation systems[20] by ensuring acceptable levels of CO_2 concentration for the health and welfare of occupants. Raised CO_2 levels have also been associated with a decrease in human performance, especially in schools, where there is usually high occupancy and low ventilation rates, and the main pollution source is the occupants (e.g.,)[21,22]. At normal levels, the CO_2 presence has no measurable adverse effects on body, but if breathing is compromised or exposed to large amounts of this gas, a wide range of side effects can occur (e.g. rapid breathing, rapid heart rate, and fatigue).

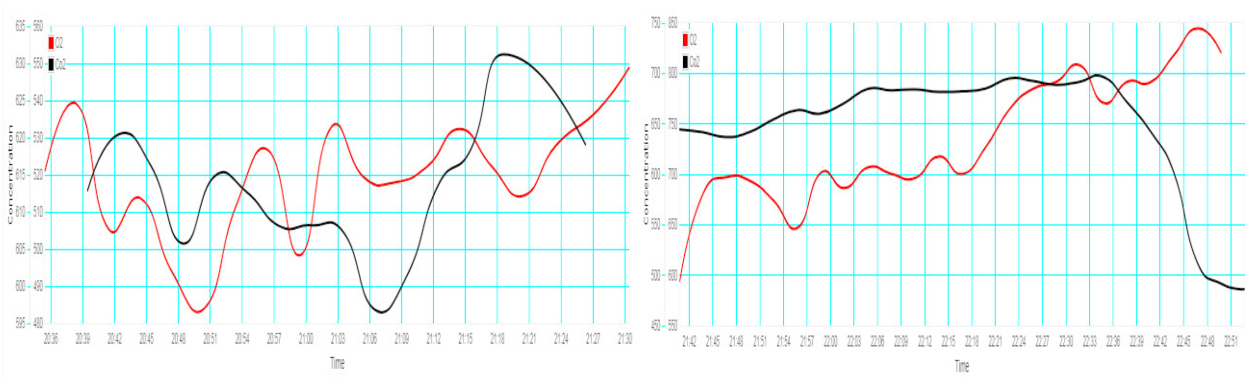
Recent studies showed also that maintaining the precise balance of oxygen-saturated blood is vital to the person's health. Peripheral capillary oxygen saturation (SpO_2) [13] is an estimate of the amount of oxygen in the blood. SpO_2

can be measured by pulse oximetry. It works by emitting and then absorbing a light wave passing through blood vessels (or capillaries) in the fingertip. A variation of the light wave passing through the finger will give the values of the SpO_2 measurement. However, we are using a pulse sensor that measures the variation of O_2 concentration instead of SpO_2. But, there is a direct relationship between CO_2, O_2 concentration in the blood, and SpO_2. According to (Cruickshank & Hirschauer, 2004) [22], as the indoor CO_2 concentration increases, the partial CO_2 in the blood increases (i.e., less O_2 concentration), which will decreases SpO_2.

In order to test the correlation between CO_2 and O_2 concentration in the blood, we have realized a scenario using the platform. The scenario is composed of two main components as follows:

- **Kaa applications:** used to collect data from two different entities: from the environment (i.e., security staff office) and its occupants. We created two Kaa applications, one to gather environmental data, and another one to gather the occupant's vital signs. Both applications transmit data to the cluster for processing.
- **Storm application:** a storm processing application based on a specialized algorithm, which is represented by the Storm topology that has two spouts that read tuples from Flume then transmit them to bolts as a stream. Each bolt consumes the stream and performs the necessary logic already implemented on it.

As a preliminary experiment, we choose a security staff office with about four occupants on average. We found this office to be an ideal place to the experiment; we had CO_2 sensors, temperature and humidity sensors, and pulse sensors. Each experiment lasted about one hour, and the preliminary results were promising, even if the number of stored samples was not huge. Two subjects were participating to the experiments during two hours. Each subject took part of the experiments for almost one hour. Figure 2 shows the CO_2 and the variation of O_2 concentration values. The two figures, Fig. 2 (a) and (b), show CO_2 reading values at left side of the Y-axis and the variation of O_2 concentration at its right side. The resulting levels of CO_2 in the office originated from two different sources: brought from outdoor air and emitted by the occupants. The office was closed (with intermittent opening by the office occupants) and the ventilation system was deactivated. Real time measurements of the CO_2 levels were made to monitor the variation of concentration during the experiment.



(a) (b)

Fig. 2. CO_2 and O_2 relationships for two subjects: a) subject 1, b) subject 2

This figure shows the variation of O_2 concentration of the two subjects in terms of CO_2 concentration. The variation of O_2 concentration varies as well as CO_2 concentration changes its value. As expected, the greater the CO_2 concentration, the lower is the variation of O_2 concentration in the blood. Furthermore, as expected (see e.g., (Cruickshank & Hirschauer, 2004) [22]), as time elapses, we noticed a direct relationship between these entities. In fact, when the indoor CO_2 concentration is increased, the variation of O_2 concentration is decreased.

5. Conclusions and perspectives

In this paper, we have introduced a platform design for real-time data stream monitoring and processing. A first prototype has been developed and tested, using a small-scale healthcare scenario, in order to show the usefulness of the holistic platform in developing context-aware applications. While these preliminary results are still promising, more sensors will be integrated to show the evident relationships between these entities and occupant's behavior, e.g., if less oxygen is available to breathe, symptoms such as rapid breathing, rapid heart rate, and fatigue can result. Our ongoing work focuses mainly on extensive testing of the platform using other healthcare scenarios (related e.g., to hypoglycemia)[1]. Future work concerns its integration with the work we have done for energy efficiency buildings and smart grids in order to show the real-time data processing for context-driven controls[20].

Acknowledgements

This work is mainly supported by the VLIR-UOS funding agency, under the SELFSERV project (2016-2018), and partially supported by MIGRID project (Grant 5-398, USAID, 2017-2019) and CASANET project (CNRST-MESRSFC, 2016-2018).

References

1. V. De Florio, M. Bakhouya, D. Eloudghiri, C. Blondia: Towards a Smarter organization for a Self-servicing Society, DSAI 2016, pp. 254-260, December 1-3, 2016 - UTAD, Vila Real, Portugal
2. "Oracle: Big Data for the Enterprise", white paper, Oracle Corp., 2013.
3. S. Shahrivari and S. Jalili, "Beyond batch processing: towards real-time and streaming big data". Computers. 2014;3(4):117–29.
4. D. Sun, G. Zhang, W. Zheng, and K. Li, "Key Technologies for Big Data Stream Computing", in Big Data - Algorithms, Analytics, and Applications, pp. 193-2014, 2015.
5. J. Turner, "Hadoop: What it is, how it works, and what it can do", 2013.
6. ATZORI, L. et al. The Internet of Things: A Survey. Computer networks, Elsevier, v. 54, n. 15, p. 2787–2805, 2010.
7. HEINZE, T. et al. Tutorial: Cloud-based Data Stream Processing. 2014.
8. STORM, A. Understanding the Parallelism of a Storm Topology. 2015. Disponível em: <http://storm.apache.org/documentation/Understanding-the-parallelism-of-a-Storm-topology>.
9. STORM Concepts : http://storm.apache.org/releases/2.0.0-SNAPSHOT/Concepts.html
10. https://www.kaaproject.org/
11. S. Shahrivari and S. Jalili, "Beyond batch processing: towards real-time and streaming big data". Computers. 2014;3(4):117–29.
12. Mauna Loa Observatory, Hawaii (NOAA-ESRL) Preliminary data released May 5, 2017
13. https://help.withings.com/hc/en-us/articles/201494667-What-does-SpO2-mean-What-is-a-normal-SpO2-level-
14. F. Lachhab, M. Bakhouya, R. Ouladsine, M. Essaaidi, Performance Evaluation of CEP Engines for Stream Data Processing, in Cloudtech 2016, Marrakech Morocco.
15. D. F. Barbieri, D. Braga, S. Ceri, E. D. VALLE, and M. Grossniklaus, M, "C-sparql: a continuous query language for rdf data streams". International Journal of Semantic Computing, 4(01), 3-25.2010.
16. D. Anicic, P. Fodor, S. Rudolph, and N. Stojanovic, "A rule-based language for complex event processing and reasoning". In Conference on Web Reasoning and Rule Systems (RR 2010), 2010.
17. D. Le-Phuoc, M. Dao-Tran, JX. Parreira, M. Hauswirth, "A native and adaptive approach for unified processing of linked streams and linked data. In: Proceedings of the 10th ISWC—Volume Part I, ISWC'11. Springer, Berlin, pp. 370–388.2011.
18. D. Le-Phuoc, M. Dao-Tran, JX. Parreira, M. Hauswirth, "A native and adaptive approach for unified processing of linked streams and linked data. In: Proceedings of the 10th ISWC—Volume Part I, ISWC'11. Springer, Berlin, pp. 370–388.2011.
19. J. Gubbi , R. Buyya , S. Marusic, M. Palaniswami, nternet of Things (IoT): A vision, architectural elements, and future directions, Future Generation Computer Systems 29 (2013) 1645–1660.
20. F. Lachhab, M. Bakhouya, R. Ouladsine, M. Essaaidi, Monitoring and Controlling Buildings Indoor Air Quality Using WSN-based technologies, in the 4th International Conference on Control, Decision and Information Technologies, April 5-7, 2017, Barcelona.
21. Coley, D.A., Greeves, R. and Saxby, B.K.The effect of low ventilation rates on the cognitive function of a primary school class, Int. J. of Vent, 6, 107-112. 2004
22. S. Cruickshank and N. Hirschauer, The alveolar gas equation, Contin Educ Anaesth Crit Care Pain (2004) 4 (1): 24-27.
23. The open IoT platform with MATLAB analytics <https://thingspeak.com/>
24. Machine-to-machine (M2M)/"Internet of Things" connectivity protocol <http://mqtt.org/>
25. Flexible architecture based on streaming data flows <https://flume.apache.org>