Big Telemetry Data Processing in the Scope of Modern Internet of Things

Anton R. Fakhrazeev, Alexey Yu. Rolich, and Leonid S. Voskov

***Abstract*—This paper reviews modern ways of data preparation, acquisition and processing in projects based on Internet of Things concept. The best arrangements are considered, including strategies and techniques of network interaction, modern methods of computing organizations in projects, ways of showing and visualizing information for better client observation and realization, as well as additional technical solutions potentially applicable to developments based on the concept of the Internet of things. Selection and integration of solutions into a solitary coordinated arrangement of data collection and processing is being carried out, the extent of use and incorporation of the anticipated framework is investigated. Consequences of the experiment results examination based on the National Instruments laboratory equipment are presented.**

***Index Terms*— Internet of Things, Big data, Protocols, MQTT, Cloud computing, Wireless communication, LPWAN, LoRa, National Instruments, Educational technology.**

I. Introduction

THIS work is devoted to the study of a modern development of the industrial Internet of Things concept. The work is gone for the improvement of information obtaining and handling framework, which usage depends on a complex of present day specialized arrangements. Among the main groups of technological solutions that are considered in this work there are the following: organization of network interaction, network protocols and computing processes organization.

The chosen research direction is relevant, since the development of the Internet of Things concept is currently going on at a very active pace. During its formation, such areas as education, industry and the economy are affected This work is gone for the education sector. Many companies consider the introduction of intelligent systems and complex developments based on the Internet of Things in the learning process, as an integral stage of the development of education. Experts at Cisco [1-2] and Zebra Technologies [3] believe that integrating "smart" solutions into the learning process will make it easier to organize and gain knowledge in a more convenient and efficient format. This can be achieved through the implementation of interactive information network systems with additional collection and processing of big data that will simplify and secure the education process, as well as include additional analytical and statistical information on the activities of students and their preferences along the learning way.

The purpose of this work is to develop a system for collecting and processing big telemetry data for the subsequent simplification of public data-dependent spheres after its integration. The potential usability of this system can be achieved in the field of education in the process of conducting remote laboratory workshops.

II. Subject Area Overview

Currently, there are many projects and systems that implement the basic concept of collecting and processing big data using network connections. But, most of them are based on the virtual-type generated data. Such data is obtained in result of largescale calculations within the framework of a single closed system. An example of such generation and processing of big data is a process of analyzing the statistics of visits to a popular Internet resource.

However, within this paper telemetry data is considered, which presupposes the presence of an ever-increasing set of finite devices that periodically send out information determined by their parameters. The total amount of stored data from the devices in this case will be considered big telemetry data, which also needs processing and additional protection and integrity providing. To ensure the necessary quality of work with such a system, it is necessary to implement it based on methods that allow sending data packets and processing them at minimal cost, regardless of the system coverage area and deployment degree.

IEEE will do the final formatting of your paper. If your paper is intended for a conference, please observe the conference page limits.

A. Method 1. Network protocol

The projected system involves the interaction of the system nodes over the Internet, which makes it important to consider the network protocol which will be used in the process of receiving and transmitting data within the system.

Most modern solutions are based on the use of standard protocols that implement the "request-response" logic. This method of organizing data transmission in the network is quite common due to its reliability and redundancy in the

The article was prepared within the framework of the Academic Fund Program at the National Research University Higher School of Economics (HSE) in 2017- 2018(grant №17-05-0017) and by the Russian Academic Excellence Project "5-100".

transmission process. However, in the scale of the system with a significant number of end nodes constantly sending telemetric data to the network implementation, this way of organizing the network interaction is much inferior to another "publisher-subscriber" method.

In the case of using the publisher-subscriber logic, MQTT protocol [4], there is no need for additional service information fields in the data packets, which greatly simplifies the delivery process and reduces the volume of each piece of information sent from the end device. This allows the telemetry data of devices to be delivered with much higher speed and less network channel load.

The overall logic of system interaction, built on the basis of MQTT protocol and “publisher-subscriber” template, consists of few steps:

1) *Broker identification*
Devices and clients connect to a common well-known broker, logical MQTT data controller, that can be an Internet service that is publicly available over the Internet. Devices offer their specific identifier, under which they are registered in the system.
2) *Broker interaction*
After registration and authorization process, devices can send telemetric data to the broker. Each packet of data includes a small amount of extra information fields, which help broker recognize sender device.
3) *Broker storage*
Broker stores received information with the addition of data tagging, to achieve structured kind of database.

Besides the interaction with the end devices, this logic provides opportunity to control data delivery to the users, via the ability to subscribe to device’s data. This allows to receive contemporary information from the system. The client in this logic is commonly understood as the 3rd party application or interface, which can be connected to the “publisher-subscriber” template. Within this research, it is considered to use such opportunity to integrate the ability to process and analyze data, using service-based applications.

The following table shows the major differences between “request-response” and “publisher-subscriber” logics.

TABLE I
“REQUEST-RESPONSE” AND “PUBLISHER-SUBSCRIBER” COMPARISON

Feature	REQUEST-RESPONSE (HTTP/HTTPS)	publisher-subscriber (MQTT)
Data exchange rate	medium	high
Traffic usage	high	low
Security and encryption	+	+
Devices direct access	+	-
Message size	big	small
Data distribution	1 to 1	1 to N
QoS levels	1	3

The MQTT protocol and “publisher-subscriber” logic is the best solution for the systems with high message transferring rate and a big amount of data-generating end devices. This makes it reasonable to implement this logic into overall data acquisition and processing system.

B. *Method 2. Cloud computing*

In modern projects, the services of cloud computing [5] are being integrated increasingly often. They are used to organize a convenient access to remote distributed computing resources. Such resources can be configured and used to provide storage and processing of information. There are currently many cloud platforms that provide the opportunity to implement IoT projects based on such platforms and services.

Among the main competing cloud platforms that support the work with services based on the IoT concept, the following should be distinguished: IBM Bluemix [6-7], Microsoft Azure, Amazon Web Service, Google Cloud Platform.

TABLE II
CLOUD PLATFORMS COMPARISON.

Feature	BLUEMIX	AZURE	AWS	GCP
IoT	+	+	+	+
Machine Learning	+	+	-	-
Big Data storage	+	+	+	+
MQTT usage	+	+	-	+

When using cloud platforms, an important problem is solved, which is in the need of a private server with additional services for storing and analyzing data implementation. While building project solution on the basis of cloud computing, it becomes possible to use modern pre-installed services on hosting that allow you to implement the required system logic in a short time and with minimal resource costs. To organize a method for storing large telemetric data and carry out their subsequent processing based on the available algorithms of machine learning.

Within this work, IBM Bluemix cloud service is used. This service represents PaaS distribution model with some IaaS capabilities. This allows to use this Platform in several ways:

1) *Broker*
Basic broker logic realization, based on “publisher-subscriber” template. IBM Bluemix provides such opportunity to use integrated IoT platform, which consists of MQTT hub, allowing to handle and manipulate device telemetric data.
2) *Storage*
Database storage, which will allow to store, and aggregate received end devices data into single big telemetric data complex.
3) *Data processing*
Data analytics service, which will allow to manage and process data, stored in the database. This includes the capabilities of machine learning and insights generation.

C. *Method 3. Modern wireless technologies*

In addition to the logic of sending data to the network, there is a need to determine the method of communication of the nodes of the system due to consideration of the network interaction within the system.

On the scale of the Internet of Things concept, it is advisable to use wireless connections of the end devices, since this allows to provide an additional coverage area, as well as flexibility and extensibility of the system. Wireless communication can be implemented in a variety of ways, including modern solutions based on Wi-Fi, LPWAN and other technologies. LPWAN technology in the modern world has the widest distribution in projects based on the Internet of Things concept, so in the course of this work it was decided to consider the technological solutions directly related to this technology. LPWAN technology is a new approach in radio communication and is used for devices in large networks. The main feature of this technology is its low power consumption and wide coverage (in the territorial sense). The basis of the principle of data transmission is the well-known dependence in modern radio systems - for each bit of encoded transmitted information sufficient power is allocated so that the data can be separated from the general network noise. Consequently, a decrease in the data transfer rate entails an increase in the coverage of the system. Unlike other wireless data transfer technologies such as ZigBee or LTE, the LPWAN is characterized by its energy efficiency and range, sacrificing the data transfer rate. However, since it is supposed to use sensory data and information from sensors or remote systems, the amount of transmitted data is minimal, so the speed can be neglected.

The following list represents the most common advantages and disadvantages of using LPWAN technology:

1) Large coverage area, which can be up to 15 kilometers.
2) Low energy consumption in any device, used in the system
3) Ability to use unlicensed frequencies, including Russia.
4) High signal penetration rate.
5) Low speed and bandwidth of data transferring.

The above features of this technology make it possible to implement distributed systems consisting of a set of finite node-devices located at great distances from one another, which periodically must send telemetry and their state data to the network for further analysis and processing.

A particular case of LPWAN wireless solution is LoRa (LoRaWAN) [8] technology. This technology involves the transfer of small amounts of data from each of the end devices over long distances with minimal energy costs. This way of system operating allows collecting a large amount of telemetry information from a variety of devices, while maintaining the optimum level of information transfer speed for small data of each device. In addition to LoRa, several other technologies have also been built based on LPWAN.

TABLE III
LPWAN STANDARDS COMPARISON.

Feature	LoRa	Sigfox	STRIZH (rus)
Frequency	868 MHz	868 MHz	868.8/433 MHz
Range	5/45 km	10/50 km	< 40 km
Scalability	high	low	high
Speed	< 50 kbs	< 1000 kbs	50 bs

Among the highlighted technologies, LoRa has the most effective combination of features, so it is advisable to use this wireless technology in the scale of projects based on the Internet of Things.

III. EXPERIMENT RESULTS

In general, this system is divided into 3 key types of end nodes: device-data generators, concentrators or gateways and a cloud service. Generator devices send telemetry data to the hubs via radio network, which receive a packet of data over the radio channel, form a new packet based on the MQTT protocol and send it to the broker that is located on the cloud platform. The cloud platform processes each of the received data blocks, forms a database that represents a large amount of telemetric information and subsequently performs additional processing and analysis of the resulting telemetry information store from the system devices.

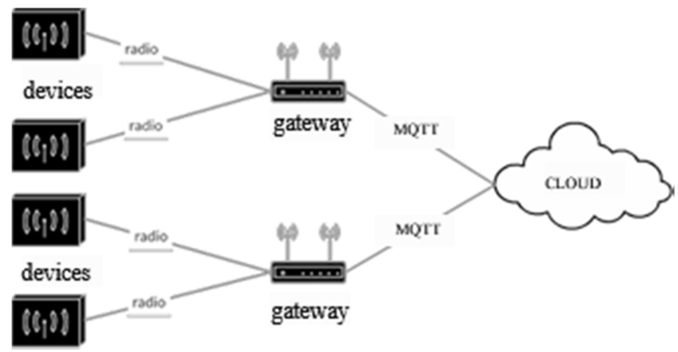

Fig. 1. Developed system structure for end-point devices, sharing messages through gateways via wireless protocols.

To demonstrate the efficiency and effectiveness of the complex of technological solutions, National Instruments equipment was used. With this equipment, the basic logic of end nodes interaction was built and the delay in sending and receiving data was analyzed, which is one of the most important aspects of the real-time system operation. On the scale of the interaction, it was decided to exclude one of the methods that involves the use of wireless LPWAN technology due to the lack of implementation capabilities based on equipment from National Instruments.

A statistical information on the average delay values and data loss in the process of receiving and transmitting messages depending on the intensity of sending is shown below.

TABLE IV
INTENSITY, DELAY AND DATA LOSS DEPENDENCY

data/min, intensity	%, DATA LOSS	ms, delay
1	0	154
10	0	160
30	4	180
60	6	200
75	9	210
120	50	230

As can be seen from the statistics, this system can be used in working with equipment that does not involve loads of using a data channel to send a message more often than once

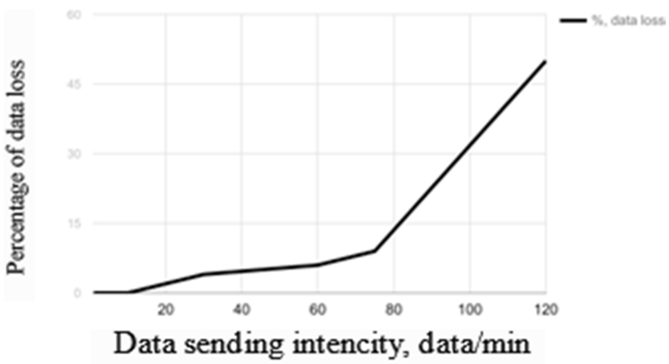

Fig. 2. Data loss dependency chart. Horizontally – data sent in a minute, vertically – percent of data loss due to technological reasons.

per second. This allows to conclude that this system can be used in many areas where the frequency of delivery and updating of telemetric information does not affect the system performance.

IV. Conclusion

This data collection and processing system has the simplest structure and flexibility, which allows to connect an unlimited number of telemetry information generators, as well as other devices that can be potentially used as a source of useful data.

From the main areas of integration of this system, the education scope can be identified. The system can be used as a method of organizing remote laboratory workshops with additional collection of statistical information. All these data will form big data, which contains information on the progress of the workshops implementation, telemetry data on the work of laboratory stands and statistical information on the activities of students and their preferences, based on an analysis of their actions during interaction with the system.

Practical significance of this project and system implementation consists of the following aspects and capabilities:

1) In education, it is implied to increase the efficiency of laboratory workshops execution, based on special equipment of companies which provide opportunities to connect their devices over the Internet.
2) Simplification of telemetry data acquisition and processing even within the large territorial range of device scatter.
3) Providing basics knowledge of contemporary IoT technologies existence.

References

[1] Michelle S., Ana S., Jim B. Cisco research - Education and the Internet of Everything [Online]. Available: http://www.cisco.com/c/dam/en_us/solutions/industries/docs/education/education_int ernet.pdf

[2] Cisco – The Internet of Things Reference Model. [Online]. Available: http://cdn.iotwf.com/resources/71/IoT_Reference_Model_White_Paper_June_4_2014.pdf

[3] Zebra Technologies – How the Internet of Things Transforming Education [Online]. Available: http://www.zatar.com/sites/default/files/content/resources/Zebra_Education- Profile.pdf

[4] Hunkeler, Urs, Hong Linh Truong, and Andy Stanford-Clark. "MQTT-S—A publish/subscribe protocol for Wireless Sensor Networks." Communication systems software and middleware and workshops, 2008. comsware 2008. 3rd international conference on. IEEE, 2008.

[5] Reese, George. Cloud application architectures: building applications and infrastructure in the cloud. " O'Reilly Media, Inc.", 2009.

[6] Lampkin, Valerie, et al. Building smarter planet solutions with mqtt and ibm websphere mq telemetry. IBM Redbooks, 2012.

[7] Kobylinski, Kris, et al. "Enterprise application development in the cloud with IBM Bluemix." Proceedings of 24th Annual International Conference on Computer Science and Software Engineering. IBM Corp., 2014.

[8] Adelantado, Ferran, et al. "Understanding the limits of LoRaWAN." arXiv preprint arXiv:1607.08011 (2016).W.-K. Chen, *Linear Networks and Systems*. Belmont, CA, USA: Wadsworth, 1993, pp. 123–135.