IOT BASED CLUSTER FOR DATA COLLECTION AND ANALYSIS

Žarko Bogićević[1]; Milorad Murić[2];
[1] Technical school, Užice, Serbia, zarko1993@hotmail.com
[2] High school of vocational studies, Užice, Serbia, muricmilorad@gmail.com

Abstract: *The main goal of this paper is to analyze an IOT based cluster infrastructure for collecting specific data sets. The idea is to create a model and IOT based device that either sends data to a specific server or a cloud based server and uses it for further analysis. This model includes a practical example that is being piloted on an airport and collects weather data relevant to the airport and its operations. An IoT ecosystem consists of web-enabled smart devices that use embedded processors, sensors and communication hardware to collect, send and act on data they acquire from their environments. IoT devices share the sensor data they collect by connecting to an IoT gateway or other edge device where data is either sent to the cloud to be analyzed or analyzed locally. Sometimes, these devices communicate with other related devices and act on the information they get from one another. Another goal of this project is to create an affordable solution that can be used in small and medium scale clusters. This paper aims to showcase an affordable cluster that can be developed for any use case scenario, as more and more services switch towards the cloud, analyzing data and making predictions based on those analysis is crucial for success.*

Keywords*: IOT, cluster, data collection, data analysis*

1. INTRODUCTION

If we look at devices in the world around us, whether they are remote control devices, robots, coffee machines, home bakers, traffic signals, it can be noticed that in each of these devices there is a block that controls its work. Nowadays, managing the operation of various devices in a human environment is the easiest way to realize using computers. One of the limitations of this approach from the historical point of view was the size of computers and the consumption of electricity. The development of semiconductor technology created conditions for us to have computers at the level of integrated circuits in the form of microprocessors and microcontrollers. From the aspect of managing modern devices and systems of special importance microcontrollers play a huge role.

In this paper a solution to building small scale IOT based cluster in order to gather data and create specific data sets. Several solutions will be documented within this paper, depending on the infrastructure and resources available. The data collected is relevant to the airport itself, however, the design is modular and it may be implemented elsewhere. The solution provided must fulfill several requirements:

- Low cost;
- Reliable;
- Simple to use;
- Easily programmable;
- Universal.

This device must be low cost so that many of these devices could be purchased in order to create clusters and gather data where budgets are low. However, these devices must also be reliable and the low cost requirements must not sacrifice device reliability. Universal and standard equipment must be used in order to ensure the simplicity and programmability of these devices.

When all of the previously mentioned is taken into account, the only platform that comes to mind, which fulfills these requirements is an Arduino based platform. One of the main reasons Arduino comes to mind is it's availability, the wide array of modules available, a constantly expanding community through which users may get the help needed etc.

IOT clusters are useful for most businesses, from controlling devices, gathering data for analysis for business critical domains, while, in the airport example, the data could be safety critical, such as ice on the runway, poor flying conditions etc. Many more things could be determined and predicted if an IOT data cluster has been applied and used properly. There are systems available already, but most of them are not affordable or those that are do not provide proper reliability.

2. THE ARDUINO PLATFORM

Arduino is an open-source electronics platform based on easy-to-use hardware and software. Arduino boards are able to read inputs - light on a sensor, a finger on a button, or a Twitter message - and turn it into an output - activating a motor, turning on an LED, publishing something online. You can tell your board what to do by sending a set of instructions to the microcontroller on the board. To do so you use the Arduino programming language (based on Wiring), and the Arduino Software (IDE), based on Processing [1].

Due to the way Arduino works and it’s simple and accessible user experience, it is used in thousands of different projects and applications. The Arduino software is easy to use for beginners, however, it is also flexible enough for advanced users.

The supported operating systems for the Arduino software are:

- Windows
- Mac
- Linux

Also, there is an Arduino cloud software which works through the browser itself and communicates with the boards using a browser plugin that needs to be installed. The projects and libraries are stored online which makes it easier to move projects around, instead of all the project files being saved offline on a single computer.

Arduino is widely used in classrooms as well. Teachers and students use them to build low cost scientific instruments, to prove chemistry and physics principles, or to learn basic programming and robotics. It may also be used for building interactive prototypes, experiments with instruments and so on. One of the main advantages Arduino offers is its ease of use even for non-tech people.

The hardware used for this paper an Arduino Uno microcontroller as shown in the picture below.

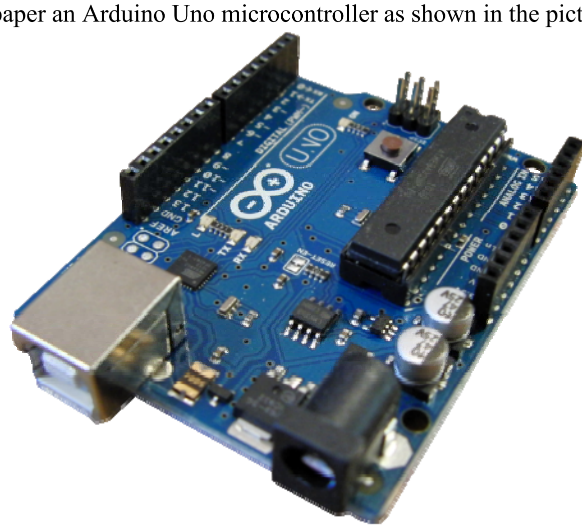

Figure 1: Arduino Uno microcontroller

Another Arduino based microcontroller that is excellent and is used in the paper is the Arduino Nano microcontroller.

Figure 2: Arduino Nano microcontroller

These microcontrollers enable the creation of devices using a combination of hardware and software. By using an array of sensors in a combination with one of these microcontrollers a new device may be created with a specific purpose. The entire hardware and software are open source, which means any modifications can be made or even smaller microcontrollers may be created using the provided documentation. Some of the modules available include WIFI, NFC, LCD screen, LED dot matrix, RFID, SD, GPS, GSM modules etc. There are also other Arduino versions available such as Due, Leonardo, Mega 2560, Mini etc.

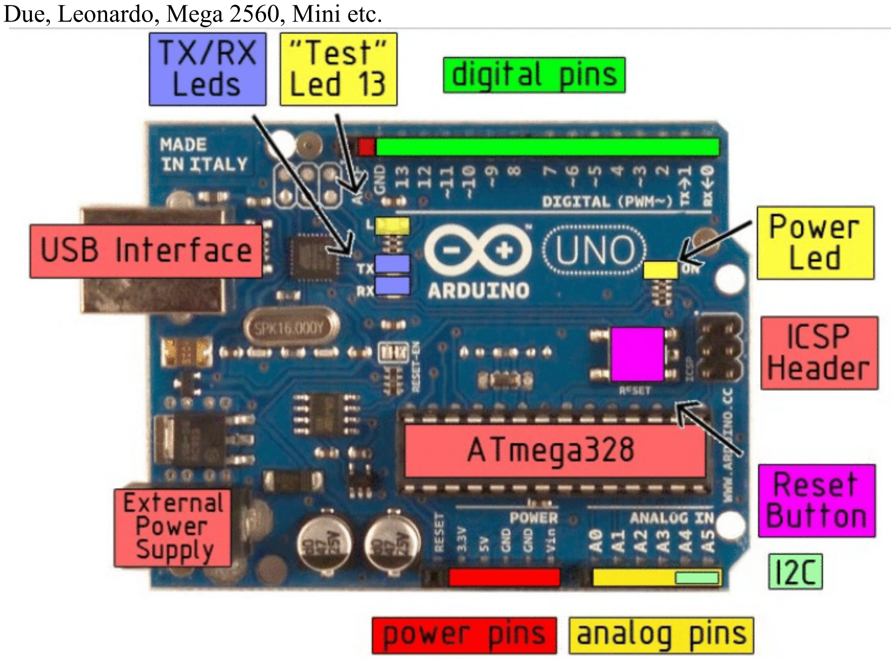

Figure 3: Arduino Uno block diagram [2]

As shown in the previous image, Arduino is capable of powering other modules with 3.3v or 5v. There are pins for analog and digital sensors as well. The USB interface can be used for communication and programming, but it can also be used to power the Arduino boards. There are four LEDs on the board. One of which shows the ON signal indicating that the board has power delivered to it. RX and TX blink when the board is receiving and sending data thru the USB port. And the last LED is connected to the digital pin 13 and is used for testing purposes only, programming the PIN 13 to blink is used to check if the board is working properly.

Another advantage of the Arduino platform is the ability to expand. Arduino possesses modules which can be stacked upon the microcontroller board. These modules are called Arduino shields.

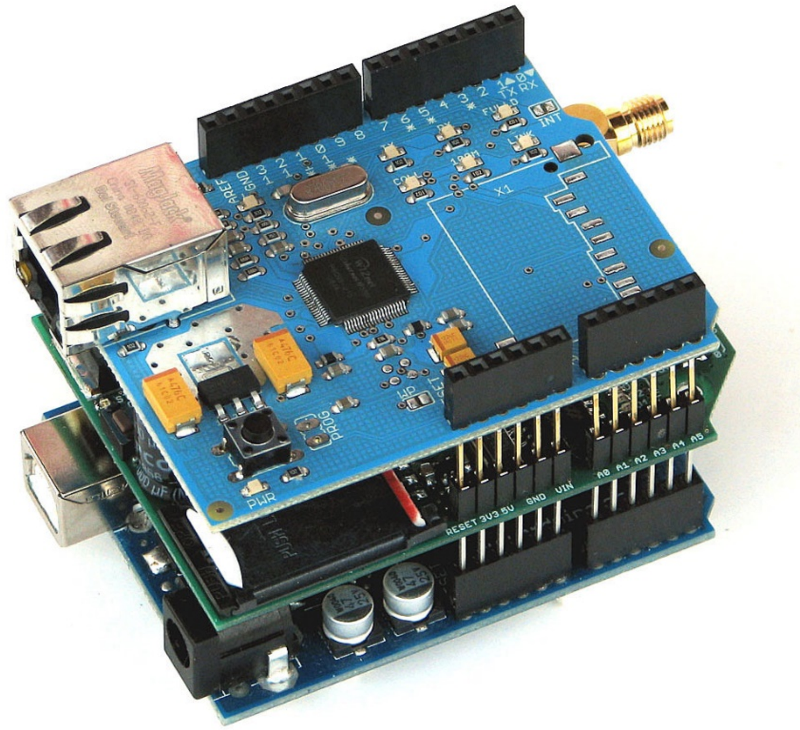

Figure 4: Arduino Shields

3. THE RECOMMENDED SOLUTIONS

There are several ways these devices could be made. Depending on the location of the devices themselves, whether they have access to LAN, WIFI, GSM or no access to the Internet at all. In cases where there is Internet access in one form or another, the devices could connect to a remote server and dump the data gathered there, they could act as a server themselves and data could be accessed by connecting to each device, but these solutions are not practical enough and require a lot of work for something that could be automated. In the case where the devices have Internet access the best solution is to write a script which will be used to transfer the data automatically into a database. Databases are ideal for this sort of data as the data could be further manipulated and analyzed much easier if it is already sorted in a database.

Since the previously mentioned goal of this paper is to provide a low budget and reliable solution. One solution could be a MySQL database hosted on a local PC to which the devices would periodically connect and insert the data gathered. It is best to have a dedicated PC or server for this purpose, however, keeping in mind the low budget nature of this paper, a Raspberry Pi would work fine on a smaller scale cluster or a larger cluster with optimized times of insertion. Another thing to keep in mind, when working with a Raspberry Pi or a similar single-board computer, is to have external memory via USB or another memory type. SD cards which are used will fail or corrupt a lot faster if the database is stored on them, as it writes in the same sectors which is not ideal for SD cards.

There are found solutions recommended in this paper:

1. An Arduino based microcontroller collecting data from a sensor array wired to it. The data is sent via USB or LAN using an ethernet shield.
2. An Arduino based microcontroller collecting data from a sensor array wired to it. The data is sent via WIFI.
3. An Arduino based microcontroller collecting data from a sensor array wired to it. The data is sent via a GSM module by either sending out text messages or using an Internet connection provided by the GSM module.
4. An Arduino based microcontroller collecting data from a sensor array wired to it. The data is collected locally to an SD card.

This paper will cover two of these solutions.

The first solution will showcase a system where the infrastructure already exists, power and Internet is already available via cable connections and the device relies on an external power supply and the network cables already provided.

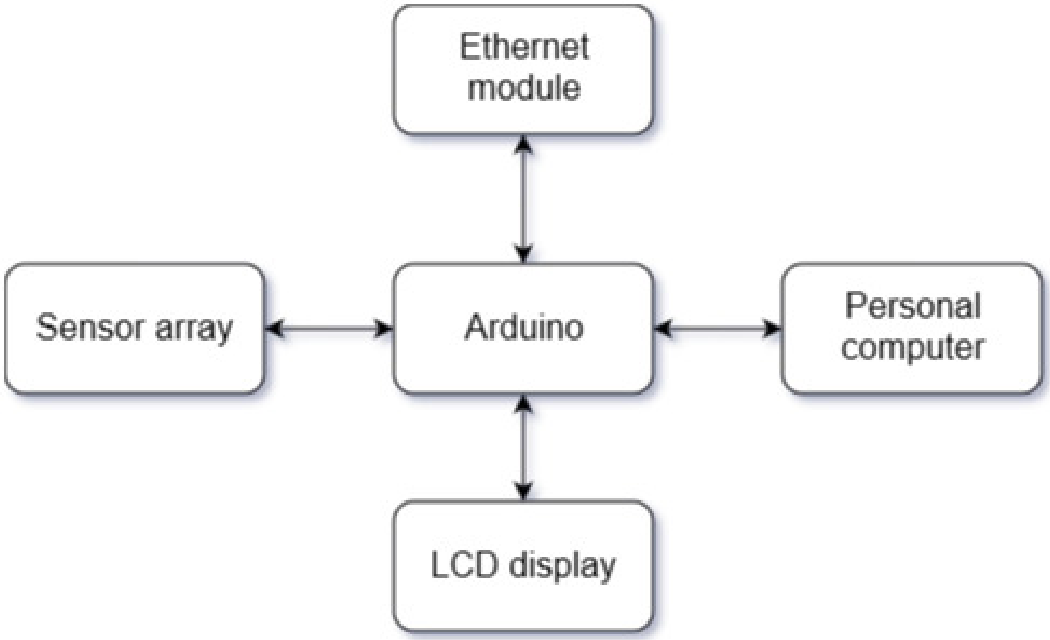

Figure 5: Block diagram of the 1st device

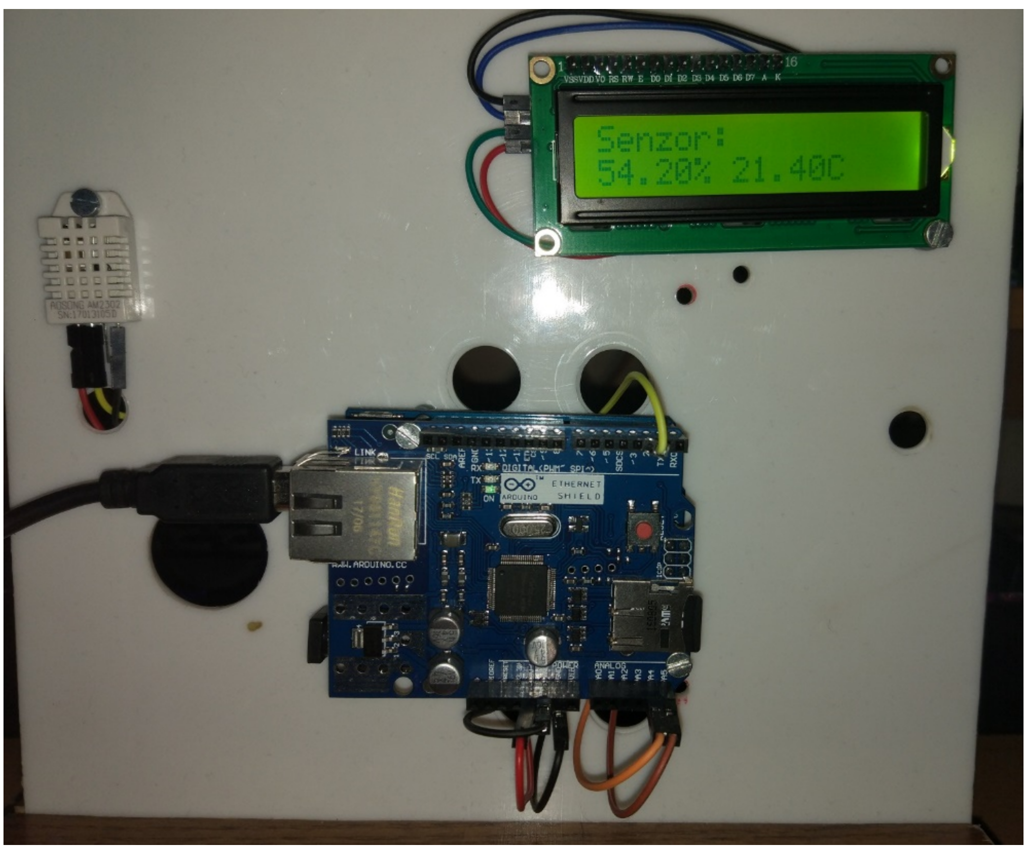

Figure 6: Demo version of the 1st device

As seen from the picture above, the 1st device requires a constant power supply, it has the ability to send data over LAN or to store it on the SD card. There is also an LCD display showing the information from the sensor. In this case the data could be sent over the network constantly as the power supply is not an issue or it may be stored on the SD card and sent periodically in bulk.

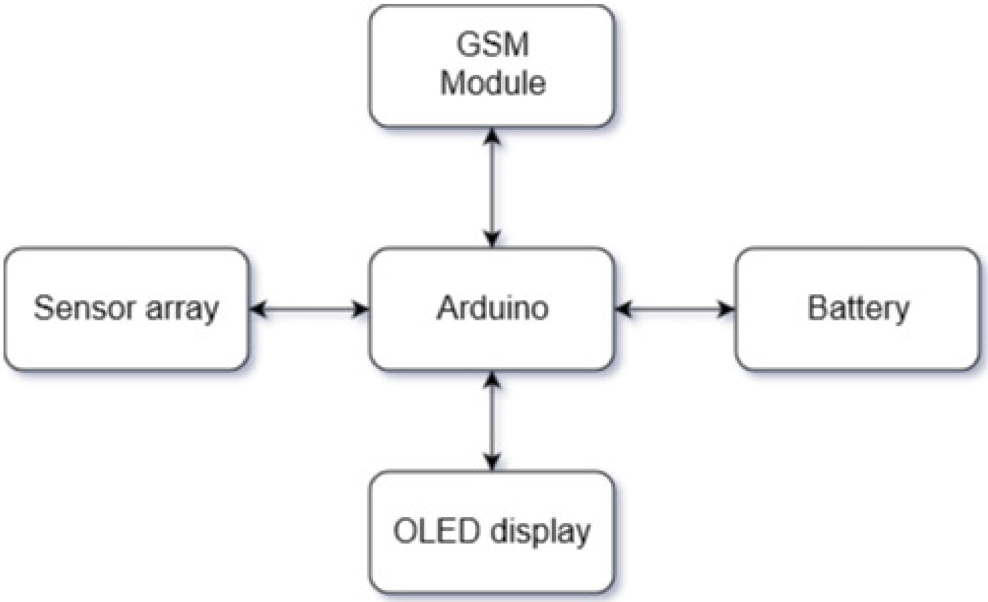

Figure 7: Block diagram of the 2nd device

The second device uses the same principal in terms of gathering data, however, it is more robust and independent, as it relies on battery power and a GSM module with an antenna to communicate, so in those terms it may be dispatched at remote location where an infrastructure does not exist or at locations which are not accessible for long periods of time.

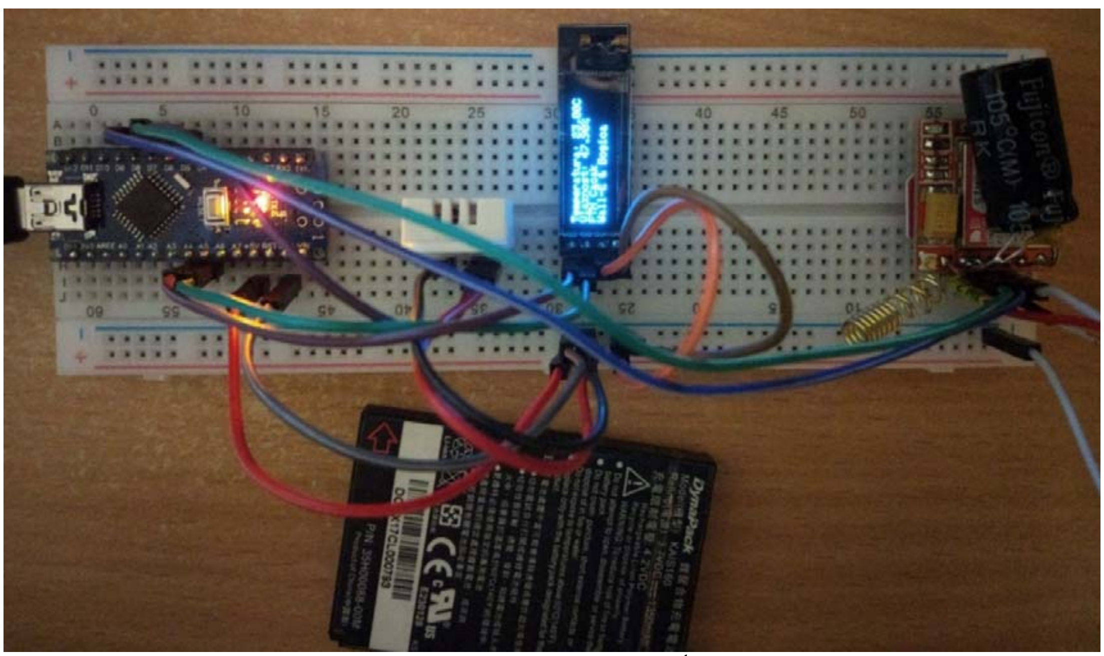

Figure 8: Demo version of the 2nd device

Depending on the size of the battery provided, this device could be powered on for months. The OLED display is used as a status check and consumes less power than a traditional LCD display. When coding a device such as this, it is important to properly code sleep intervals. During those intervals the device should go into a low power mode and preserve the battery, only waking up when it collects data and pushes it to the server over the air using the GSM module.

There is also the option of waking the device using a text message and getting data back from it as a text message reply. And both devices support USB serial communication.

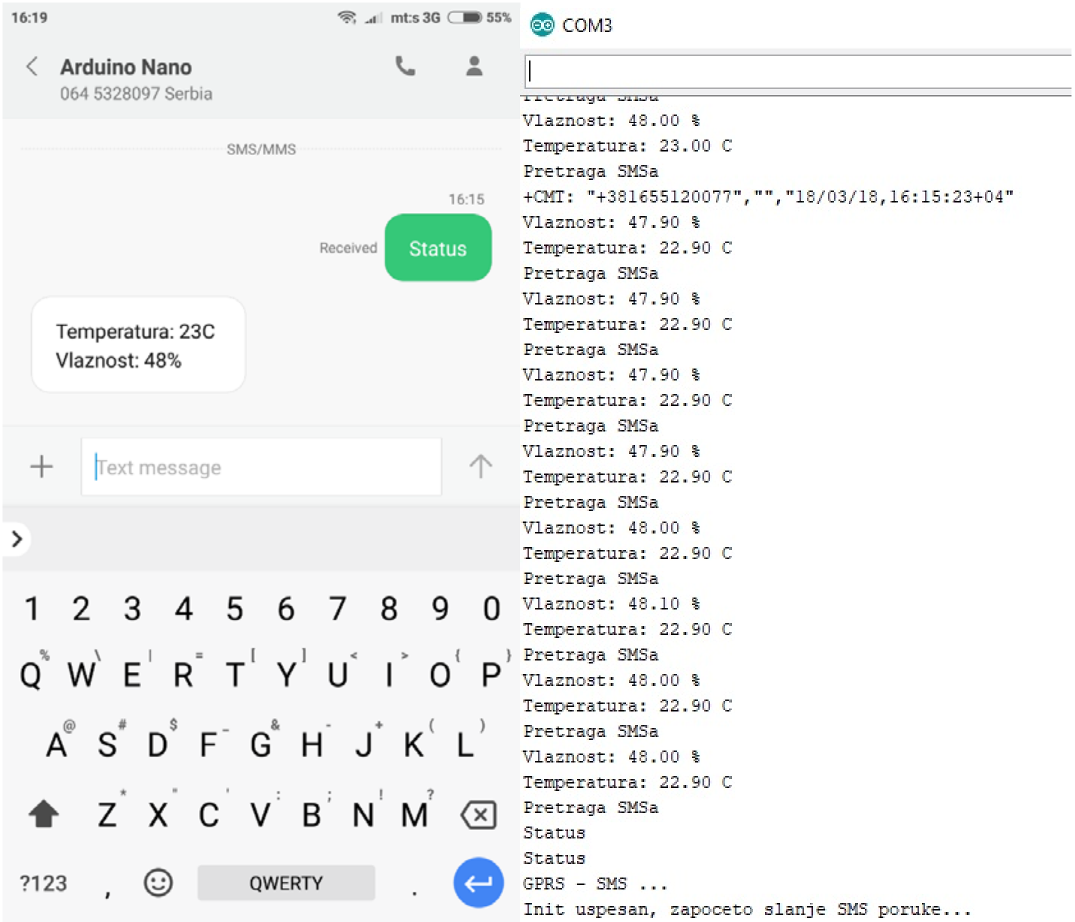

Figure 9: Communication using text messages and serial communication

4. CONCLUSION

Apart from what is shown as the demo devices, these concepts can be further improved and used for specific needs, from live monitoring and reporting to data gathering, creating specific data sets and data mining them for patterns in order to generate new information. The previously shown devices support additional sensors and with the addition of a BME, light and rain sensor can provide data relevant to the operation of an airfield. This is being piloted at the time of writing this paper and the data will be collected and compared and used to predict weather conditions in this time, next year. Such devices could be used in monitoring highway conditions, predicting fog, poor weather, heavy rain and warning drivers approaching those areas. The universal part of these devices comes into play here, they can be used for many different purposes with little to no modification from the previously shown general approach..

In a world connected, from the industry, housing, agriculture, healthcare, retail, supply chains, businesses etc. it is just a matter of time before every aspect of our lives will be connected in some way. There is a major reason for this, connecting, data gathering and analyzing allows us to determine why something is happening, how it is happening and what we could do to fix or prevent it. In such a world, small scale affordable clusters, like the one demonstrated in this paper will be implemented in non mission critical systems. Simply put, these devices improve our daily lives and with the data gathered could potentially provide a business advantage and a higher quality of life in general.

REFERENCES

[1] https://www.arduino.cc/en/Guide/Introduction (25.03.2019)

[2] https://www.researchgate.net/figure/Block-diagram-of-the-Arduino-Board-The-Basic-Features-of-Arduino-Board-BULLET-The_fig3_307477633 (25.03.2019)