Int. J. Advance Soft Compu. Appl, Vol. 9, No. 3, Nov 2017
ISSN 2074-8523

Real time Photovoltaic Data Collection via IoTs and Analytics on Hadoop Cluster

Zainal H. C. Soh, Shafiq N. Kadir, Syahrul A. C. Abdullah, and Mohammad N. Ibrahim

Faculty of Electrical Engineering, UiTM Pulau Pinang, Penang, Malaysia
zainal872@ppinang.uitm.edu.my, shafiqfezi@yahoo.com,
mnizam@ppinang.uitm.edu.my
Faculty of Electrical Engineering, UiTM Shah Alam, Selangor, Malaysia
bekabox181343@salam.uitm.edu.my

Abstract

The project presents a prototype of real time sensor data collection and analytic of Photovoltaic (PV) solar panel via Internet of Things (IoTs) at IoTs Foundation on IBM Bluemix Cloud Platform. The data collection involves collecting and storing of the light sensor data and the photovoltaic voltage data of solar panel from Intel Galileo Sensor Node to IoTs gateway at IoTs Foundation using Message Queing Telemetry Transport (MQTT) protocol and stored on the HDFS on Hadoop Cluster. The Hadoop Cluster is leased from the IBM Bluemix Cloud Services. The stored sensor data are analysed to investigate the relationship between the photovoltaic voltage and the intensity of light during noon and dusk. The data collection rate of sensor data message is around 983 milliseconds from Intel Galileo Sensor Node to an IoTs gateway. The analysed results also show that the photovoltaic voltage is increase with higher intensity of sunlight. In conclusion, the PV data collection system proved the capability of the IoTs platform to collect the sensor data in real time environment.

Keywords: *Internet of Things (IoTs), Message Queing Telemetry Transport (MQTT), Hadoop Cluster, Photovoltaic Solar System.*

1 Introduction

Renewable energy such as solar energy is seen as crucial alternative source of energy for the future. Recently, a lot of research has been conduct on photovoltaic solar system by many scientists around the globe. In the photovoltaic solar system, many problems will be encounter since many factors can interrupt the operation of the system. The solar cells have non-linear I-V characteristic. The module itself has very low efficiency and low power output depends on solar insolation level and ambient light and temperature. The quantity of uncertainty associated with an energy estimate is affected by many factors [1].

Photovoltaic energy is influenced by a certain factor; it can by itself and environmental influence characteristic. Most significance factors are cumulative solar irradiance, power rating at standard condition, operating temperature, maximum power voltage, dependency on solar irradiance level, soiling, variation in solar spectrum, and optical losses when the light from the sun is at high angle of incidence [2]. From the case above, it shows that many factors that affect the efficiency of the solar panel is influence from the light irradiance or intensity of sunlight. To study this phenomenon, this data from the system must to be collected in real time manner. Further investigation need to be carried out in order to maximize the output power generated from the solar panel.

In order to continuously monitor and analyses the I-V characteristic data in photovoltaic solar system, a data analysis system is developed and implemented based on historical data and a real time machine log from sensor data. For the analysis of system to be useful, a huge number of sensor data need to be collected from all solar panels and others sensor across a period of time that can possibly lead toward Big Data problem, thus required a large memory resources and computing power to store and analyses the data respectively. In managing this Big Data problem, many analysts have employ Hadoop Cluster to store, analyze and process large data utilizing HDFS. In addition, a Hadoop Cluster can ensure high reliability and availability of large data set storage. Furthermore, the Hadoop Cluster has an analytic tool that can be implemented and executed parallel to efficiently process and analyses the sensor data. The unwanted problems can always occur due to faulty component or condition at the certain time and place. The distributed implementation of the sensor node can be useful to record all the sensor data more efficiently, so that the relationship between all sensor for certain time and place can be investigated.

Several data acquisition systems have been developed for use in a wide range of applications, which include measuring, acquisition and processing environmental variables [3,4] monitoring and evaluating the performance of PV solar plant systems

[5,6], monitoring and evaluating the performance of grid connected PV systems [7, 8] and etc.

Prototype of photovoltaic data collection and analytic system consist of the voltage sensor data of the solar panel and the light intensity of light sensor which are acquired via IoTs on real time environment. Furthermore, the IoTs [9] is the major breakthrough in the ICT industries nowadays to connect all machines to internet. Therefore, enable the sensor data can be access and analyze at anyplace that has internet connection in the world using PC, smartphone a tablet. The secured and efficient IoTs management is provided by a reliable IoTs Foundation to cater this large amount data. From here, the sensor data is analyses and translated into useful information by using an analytic tool system such as IBM Analytics for Hadoop on HDFS database system.

In this paper, the remaining section is about the development of PV data collection and analytics of photovoltaic solar system as described in Section 2. The result and discussion on PV data collection system performance and their data collected from the experiment are presented in Section 3, and finally the conclusion for this paper in Section 4.

2 PV Data Collection and Analytics System Development

The section presents all system components for development of real time photovoltaic data collection via Internet of Things (IoTs) on Cloud Computing. The system consist of consists of photovoltaic solar panel, light sensor and Intel Galileo as the sensor node shown in Fig. 1. The system utilized MQTT protocol in order to connect with the IoT foundation on IBM Bluemix Cloud Services. In the IBM Bluemix, there are two services used to cooperate with this system software part which is Internet of Things Foundation (IoTF) and Hadoop for Analytics, which bounded with the IBM Bluemix.

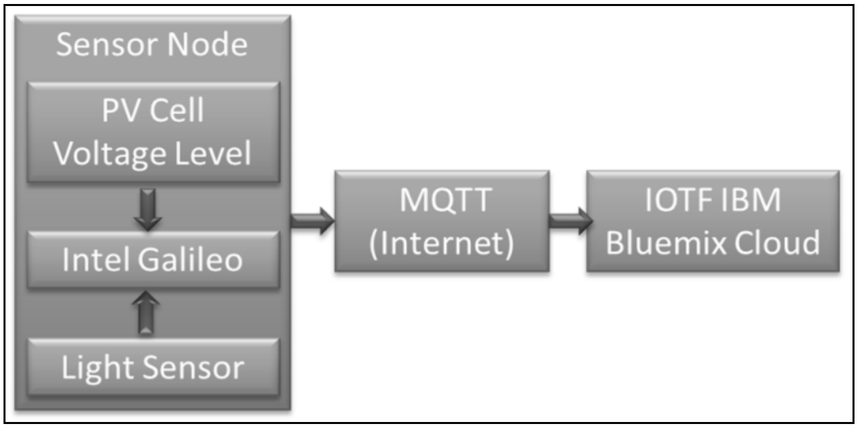

Fig. 1: Photovoltaic Data Collection System

2.1 Photovoltaic Solar Panel

From the Table 1 below, this photovoltaic solar panel is suitable for this prototyping experiment because of the small power generated from the output. The maximum current at maximum just only 0.58 A is good enough to sink into the Intel Galileo Gen 2 Development Board. The output voltage can be scale down from 20 – 21 VDC to 5 VDC by using voltage divider circuit.

Table 1: Photovoltaic Solar Panel Specification

Maximum Power (P_{MP})	10 W ± 3%
Open Circuit Voltage (V_{OC})	21.42 VDC
Short Circuit Current (I_{SC})	0.66 A
Voltage at Max. Power (V_{MP})	17.28 VDC
Current at Max. Power (I_{MP})	0.58 A
Maximum System Voltage	1000 VDC
Power Measurement at STC Irradiation 1000 W/m^2, AM 1.5, Cell Temperature 25 degree Celsius	

2.2 Light Dependent Resistor

Light Dependent Resistor (LDR) acts as analog sensor for detecting and reading the light intensity from the environment. It consists two cadmium sulphide (cdS) photoconductive cells with spectral response. The cell resistance falls with increasing light intensity [10]. Fig. 2 shows the typical lux value for certain light source, while

Table 2 is the electrical characteristic of Light Dependent Resistor (LDR). The basic operation of a LDR is when there is no light, the resistance is higher. When the intensity of light is higher, the resistance of LDR is getting lower. It is because the principle of electron is applied in this electronic part that is very sensitive with heat, friction and especially with light.

Light source	Illumination (Lux)
Moonlight	0.1
60W bulb at 1m	50
1W MES bulb at 0.1m	100
Fluorescent lighting	500
Bright sunlight	30,000

Fig. 2. Guide to Source Illumination

Table 2. Electrical Characteristic of LDR [10].

T_A = 25°C. 2854°K tungsten light source

Parameter	Conditions	Min.	Typ.	Max.	Units
Cell resistance	1000 lux 10 lux	- -	400 9	- -	Ω kΩ
Dark resistance	-	1.0	-	-	MΩ
Dark capacitance	-	-	3.5	-	pF
Rise time 1	1000 lux 10 lux	- -	2.8 18	- -	ms ms
Fall time 2	1000 lux 10 lux	- -	48 120	- -	ms ms

1. Dark to 110% R_L
2. To 10 × R_L

R_L = photocell resistance under given illumination.

2.3 Galileo Gen 2

Intel Galileo Generation 2 development board is a microcontroller based on the Intel Quark SoC X1000 application processor, a 32-bit Intel Pentium brand system on a chip (SoC). This is the 2nd generation of development board after the first generation. The design of the hardware and software pin-compatible is compatible with Arduino Uno R3 shield [11]. It supports Microsoft Windows, Mac OS, and Linux distro. The development board can also be programmed within Arduino IDE software environment. It has several PC industry standard I/O ports and features to expand native usage and capabilities beyond the Arduino environment. It has full-sized mini-PCI Express slot, 100 Mb Ethernet port, USB host port, Micro-SD slot, USB client port, 8 Mbyte NOR Flash, and 6-pin 3.3 V USB TTL UART header.

2.4 Message Queing Telemetry Transport (MQTT)

Message Queing Telemetry Transport (MQTT), is an extremely simple lightweight messaging protocol. It is designed for constrained device and low-bandwidth, latency or unreliable network. Thus make the protocol ideal for the emerging Internet of Things (IoTs) in world of connected devices [12]. By using this protocol, the data from the sensor will easily send and published to the cloud and then to the IBM Bluemix service website.

2.5 IoT Foundation (IBM Bluemix)

Bluemix is an implementation of IBM's Open Cloud Architecture, based on Cloud Foundry, that enables user to rapidly create, deploy, and manage applications on public Cloud Services. The growing ecosystem of runtime frameworks and services can be utilized in IBM Bluemix especially on the Internet of Thing Foundation, Node-RED interface and IBM Analytic for Hadoop. In addition to providing additional frameworks and services, Bluemix provides a dashboard for users to create, view, and manage their applications and services as well as monitor their application's resource usage. The Bluemix dashboard also provides the ability to manage organizations, spaces, and user access. Bluemix provides access to a wide variety of services that can be incorporated into an application. Some of these services are delivered through Cloud Foundry. Others are delivered from IBM and third party vendors [13]. The data from the sensor node is published into Cloudant Database using the MQTT protocol through IBM Bluemix IoT Foundation. The Node-RED interface provide the link to interface to IBM Analytics for Hadoop database. The stored data is analyse in the HDFS Analytics tool service such as BigDataSheet and graphical tool that bound with this tools. Hadoop Distributed File System (HDFS) is a Java-based file system that giving a scalable and reliable storage of data [14]. HDFS is tolerant to fault and built to be used on low-end or less cost hardware. Giving high access to application data and is great for cluster that has big data sets.

2.6 Sensor Node Setup

Fig. 3 illustrates the circuit connection for sensor node within the system. In this circuit, it consists with photovoltaic solar panel and light dependent resistance (LDR) as the input of the system. The Intel Galileo is a processor to process the data and send to the cloud. At the input node of photovoltaic solar panel, a zener diode is connected together in order to maintain the voltage level of photovoltaic solar panel which is limited from 20 – 21 V only. It also prevents the overvoltage from the photovoltaic solar panel to reach into the Intel Galileo since this microprocessor cannot withstand the higher voltage than the rated voltage as stated in the microcontroller datasheet.

The analog to digital conversion formula is used to read the analog reading from the sensors constructed. The formula can be represented:

$$V_{adc} = \frac{5}{2^N - 1} \tag{1}$$

For ADC conversion of LDR, the input data from LDR is change to the voltage reading. By using the voltage divider rule at the sensor node:

$$V_o = \frac{R_L}{R_L + 10k} \times 5 \tag{2}$$

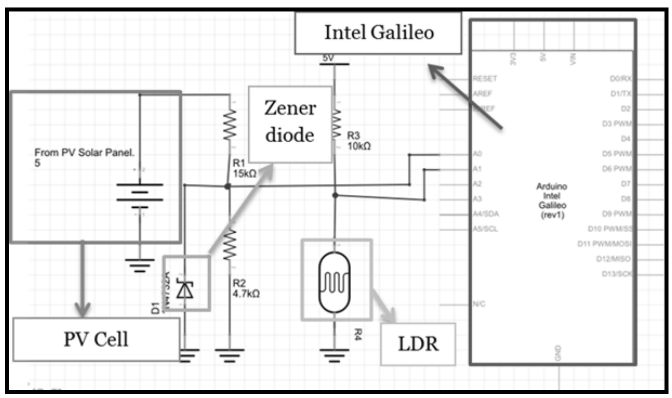

Fig. 3: Sensor Node Circuit Connection

The resistance of light dependant resistor (LDR) varies according to the amount of light that falls on it. The relationship between the resistance R_L and light intensity, Lux for a typical LDR is:

$$R_L = \frac{500}{lux} \tag{3}$$

Then,

$$Lux = \frac{\frac{2500}{V_o} - 500}{10k} \tag{4}$$

3 Result and Discussion

To evaluate the performance of photovoltaic solar panel and intensity of light, two experiments have been conduct during the noon when the sun is fully rising and during dusk when the sun is setting. The resulting message rate published to the IoT

Foundation website from the Intel Galileo is every 928 milliseconds. These showing the speed of the system almost every 1 second can be obtain from this data collection. The data tabulated in from of graph as shown in Figure 4 until Figure 7. Rapid analysis can be done by using the feature in the HDFS Analytics. The data in HDFS is process using Map and Reduce to deal with these big data.

During noon, the reading of intensity of light (lux) is high ranging from 5600 lux to 6000 lux as indicated by the Figure 7 below. This indicates during the noon, the sunlight at it maximum and the reading is high at that time. The reading of photovoltaic voltage during noon also is quite high; almost 20V can be harvests at that time. As can be seen from the graph in Figure 4 and 5, the photovoltaic voltage is declining from 19V to 18V although the light intensity is increasing from 5600 to 6000. The decrement of PV output voltage is due heat factor that caused rise in temperature of solar panel.

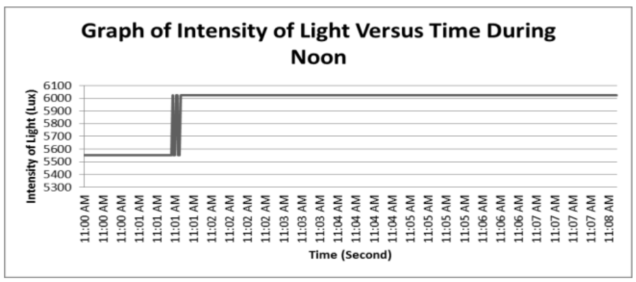

Fig. 4: Intensity of Light Versus Time During Noon

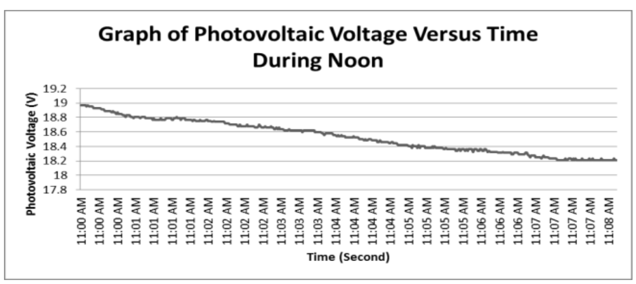

Fig. 5: Photovoltaic Voltage Versus Time During Noon

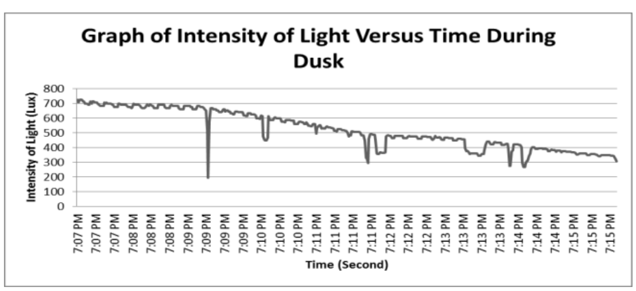

Fig. 6: Intensity of Light Versus Time During Dusk

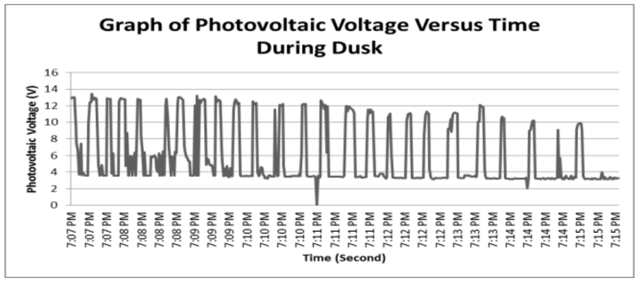

Fig. 7: Photovoltaic Voltage Versus Time During Dusk

Fig. 6 shows the intensity of light versus time during dusk. During dusk, the reading of light intensity is obviously lower than at noon, and it keeps on declining from 700 lux until 400 lux due to slight absence of the sunlight. The photovoltaic voltage reading is not very stable because the sensor does not sense the presence of light as shown in Figure 7.

4 Conclusion

In this paper, the photovoltaic data collection and analytic of photovoltaic system via IoTs on leased Hadoop Cluster is presented. The implemented data collection for Photovoltaic solar panel has its uniqueness especially in terms of their real time collection within one second transfer of sensor data from sensor node to Hadoop database. The data collection indicates that the IoTs usage is efficient with stable performance. By using HDFS Analytic tool, the data is analysed and tabulated in IBM Bluemix webpage environment. The analytic indicate the level of voltage gain from the solar panel is affected directly by the level of light intensity throughout the day. The level of gained voltage gain is decreasing as shown during the noon experiment due to excessive heat. Our future work will be on the IoTs infrastructure with the HDFS Hadoop Cluster for monitoring at the national power grid. It has potential to complement the SCADA system to monitor the national grid operation due to the system speed, stability, and reliability.

Acknowledgement

This research is partially funded by the E-Science grant – “Distributed T-Way Testing System using MapReduce Mechanism on Hadoop Cluster” from MOSTI and UiTM.

References

[1] M. Schnitzer, P. Johnson, C. Thuman, J. Freeman, a W. S. Truepower, and U. States, “Solar Input Data for Photovoltaic Performance Modeling : i,” *IEEE*, pp. 3056–3060, 2011.

[2] D. L. King, W. E. Boyson, and J. a. Kratochvil, “Analysis of factors influencing the annual energy production of photovoltaic systems,” *Photovolt. Spec. Conf. 2002. Conf. Rec. Twenty-Ninth IEEE*, pp. 1356–1361, 2002.

[3] H.E. Gad, Hisham E. Gad “Development of a new temperature data acquisition system for solar energy applications”, Renewable Energy, vol 74, 2015, pp.337-343.

[4] M. Benghanem, “A low cost wireless data acquisition system for weather station monitoring”, Renewable Energy, vol 35, 2010, pp. 862–872.

[5] Forero N, Hernandez J, Gordillo G. “Development of a monitoring system for a PV solar plant”, Energy Conversion and Management, 2006, vol 47, pp.2329-2336.

[6] Soetedjo, A.; Nakhoda, Y.I.; Suryadi, D., "Development of data acquisition system for hybrid power plant," in QiR (Quality in Research), 2013 International Conference on , vol., no., pp.197-201, 25-28 June 2013

[7] Anwari, M.; Hidayat, A.; Hamid, M.I.; Taufik, "Wireless data acquisition for photovoltaic power system," in Telecommunications Energy Conference, 2009. INTELEC 2009. 31st International , vol., no., pp.1-4, 18-22 Oct. 2009

[8] X. Zou, B. Li, Y. Zhai, and H. Liu, “Performance Monitoring and Test System for Grid-Connected Photovoltaic Systems,” in 2012 Asia-Pacific Power and Energy Engineering Conference, 2012, pp.1–4.

[9] R. Piyare, S. Park, S. Y. Maeng, S. H. Park, S. C. Oh, S. G. Choi, H. S. Choi, and S. R. Lee, “Integrating Wireless Sensor Network into Cloud services for real-time data collection,” in *International Conference on ICT Convergence*, 2013, pp. 752–756.

[10] R. Datasheet, “Light dependent resistors,” vol. 12, no. 651, p. 4, 1997.

[11] “Intel® Galileo Gen 2 Development Board.” [Online]. Available: http://www.intel.my/content/www/my/en/embedded/products/galileo/galileo-overview.html. [Accessed: 18-May-2015].

[12] “FAQ - Frequently Asked Questions | MQTT.” [Online]. Available: http://mqtt.org/faq. [Accessed: 18-May-2015].

[13] "What is IBM Bluemix?" 27-Apr-2015.

[14] "Hadoop Distributed File System (HDFS)." [Online]. Available: http://hortonworks.com/hadoop/hdfs/. [Accessed: 23-Nov-2014].