ICIC Express Letters
Part B: Applications
Volume **7**, Number **7**, July **2016**

ICIC International ©2016 ISSN 2185-2766
pp. 1475–1481

A CLOUD-BASED POWER QUALITY MEASUREMENT SYSTEM USING DISCRETE FOURIER SERIES AND ZERO-CROSSING ALGORITHM

Yeong-Chin Chen[1], Hui-Yu Shen[1,2], Hao-Yuan Chen[1]
and Chao-Hsing Hsu[2,*]

[1]Department of Computer Science and Information Engineering
Asia University
No. 500, Lioufeng Road, Wufeng, Taichung 41354, Taiwan
ycchenster@gmail.com

[2]Department of Information and Network Communications
Chienkuo Technology University
No. 1, Chiehshou North Road, Changhua City 500, Taiwan
*Corresponding author: chaohsinghsu@gmail.com

Received December 2015; accepted March 2016

Abstract. *A power quality measurement system using big data analysis approach is established to record the power usage condition. We first apply Discrete Fourier Series to obtaining the fundamental signal of the original power data because the power quality is closely related to the power signal frequency. After the fundamental sine or cosine signal is calculated, a Zero-crossing algorithm is added to compute the baseband frequency. The experiment result has demonstrated that the Discrete Fourier Series accompanied with the Zero-crossing algorithm is a valid and effective approach for baseband frequency measurement. This power quality measurement system is based on cloud computing architecture including Apache Hadoop framework and MongoDB database. The power signal information is first stored and processed in MongoDB database. The HDFS and MapReduce in Hadoop framework are then responsible for more advanced computing and analytical tasks. Additionally, the R project package is used to complete the data mining work. We finally aim to collect and calculate abnormal power events by using association rules and clustering tools in the R project package.*
Keywords: Discrete Fourier Series, Zero-crossing algorithm, Hadoop, Data mining

1. **Introduction.** Following the increasing rate of the energy usage, in addition to look for the approach for energy substitution, how to reduce and analyze the energy consumption has become more and more significant. The power signal analysis technology plays an important role for electric power and power quality [1-3]. The analysis accuracy, computing performance, and algorithm implementation do affect the overall evaluation and monitoring of the power usage condition.

The DFS (Discrete Fourier Series) approach has become most widely used in power signal analysis and metering. By way of the transformation from time domain to frequency domain, the power signal can be decomposed into many harmonics components. The FFT computation time is fast and the process operation is simple. However, if there is no enough sampled signal, i.e., large sampling interval, the signal to be estimated must become confused. To overcome this problem, the most usual solution is to try to add the sampling frequency. Once the signal frequency is increased, the more disturbance signal exists during power transmission, the more power signal frequency varies. In other words, the power signal frequency may become sensitive to the signal disturbance [4,5]. The frequency variable depends on the power usage environment such as the breakdown of the large-scale transmission system, the power loading switching, and the offline of the power supply system.

According to the power quality definition in IEEE 1159-2009 Standard [6], the power quality disturbance including voltage swelling, voltage dips, or even voltage interruption must result into the baseband signal mixed with some other harmonic signals. Accordingly, the electronic devices have inevitably been damaged under such an incorrect and variant power signal frequency. Therefore, how to measure and correct the erroneous power signal is an important and emergent topic.

Additionally, since the data collected through the advanced metering infrastructure must amount to well above 100 terabytes [7], the big data analysis for power quality measurement becomes very popular. Moreover, the cloud-based big data architecture and algorithm such as Naive Bayes classification [8,9] do affect the analysis result and performance. Hence, how to determine the software package for the cloud computing environment will be an important factor for big data analysis.

In this paper, instead of using Naive Bayes classification [8], a simple and high accuracy Zero-crossing algorithm accompanied with the Discrete Fourier Series has been provided to calculate the power signal quality. The Zero-crossing algorithm has a very simple concept and fast computing performance. However, the Zero-crossing algorithm function is apt to be constrained by large-amplitude harmonics, non-integral harmonics or even noise. Therefore, we first use the FFT approach to filter the high order harmonics and noise. Then, the Zero-crossing algorithm is applied to accurately obtaining the baseband amplitude and frequency.

This paper is divided into six sections. After the Introduction section, the Zero-crossing algorithm is provided. After that, the hardware setup and system architecture are briefly introduced. In the Data Processing Flow section, the cloud computing system architecture had been established to improve the CPU time in the distributed computing environment and to store the historical power signal into cloud database for further analytics. Finally, in the conclusion section, the overall system structure is concluded and the individual functional block is explained.

2. **Zero-Crossing Algorithm.** The Zero-crossing algorithm is an intuitional frequency detection method and fast execution without the complex mathematical formula. For a periodic signal, the period between two zero-crossing points is half the signal period. By numerical calculating for the time units of two adjacent zero-crossing points on the time axis, and finding its countdown value, we can obtain the actual frequency of the signal. Consider an electrical signal, it can be expressed as a sinusoidal function in Equation (1) and its Zero-crossing algorithm could be deduced by Figure 1.

$$y(t) = A\sin(\omega t) \tag{1}$$

where A is the amplitude, $\omega = 2\pi f$ is the angular frequency, and f is frequency of the signal. When the signal is represented in a discrete manner, and even if $t = kT_s$ is substituted, Equation (1) can be expressed by the following equation

$$y(k) = A\sin\left(\omega k T_s\right)$$

In the signal sampling process, the X axis is represented in the scale of the degree of diameter, the angular difference between successive sampling points is

$$\Delta\theta = \omega T_s = 2\pi f T_s \tag{2}$$

where T_s is the sampling time interval. Assuming a full cycle of sampling points is m, signal periods are equivalent to sampling interval multiplied by the number of points. Therefore, the signal frequency can be calculated by the following formula

$$f = \frac{1}{mT_s} \tag{3}$$

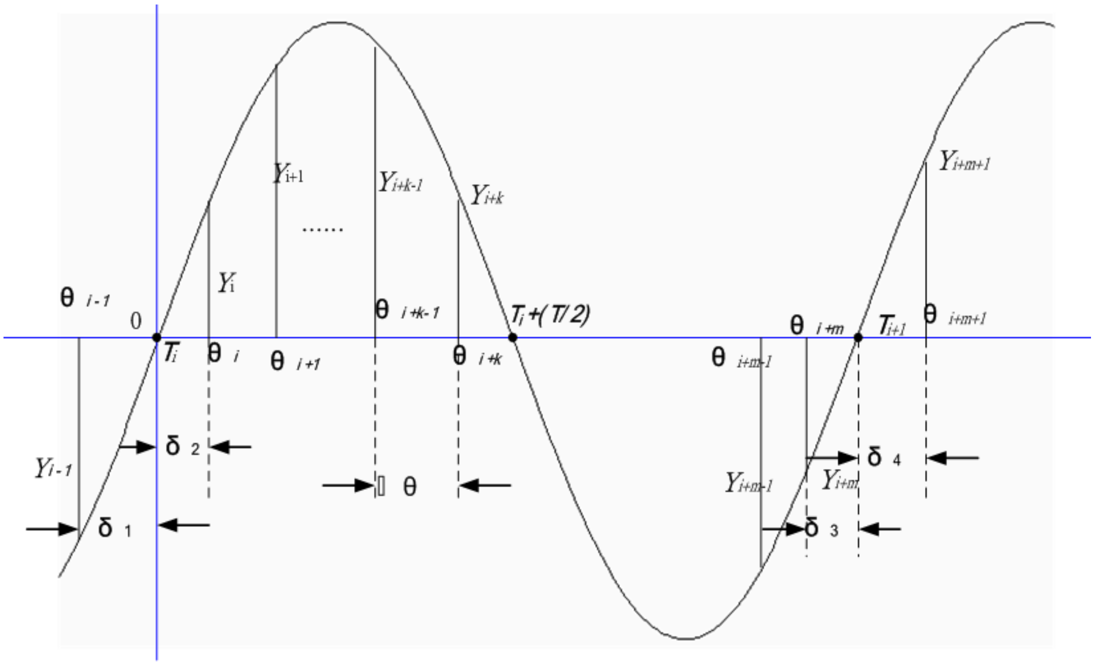


FIGURE 1. Schematic waveform for Zero-crossing algorithm

In practical applications, since the sampling frequency is not an integer multiple of the actual signal frequency, the signal period is not an integer multiple of the sampling period. Therefore, most of the exact m values are not an integer value. As we investigate the T_i period of the signal in Figure 1, there are two zero crossing points $(T_i, 0)$ and $(T_{i+1}, 0)$ with voltage magnitude transition from negative to positive in this period, where T_i is between the points (θ_{i-1}, Y_{i-1}) and (θ_i, Y_i), and T_{i+1} between (θ_{i+m}, Y_{i+m}) and $(\theta_{i+m+1}, Y_{i+m+1})$. The exact periodic of this sinusoidal wave can be expressed as

$$T = mT_s + \delta_2 + \delta_3 = m_A T_s = m_A \Delta\theta \tag{4}$$

where $T_s = \Delta\theta$ is the sampling interval. Because the signal period is not an integer multiple of the sampling interval, m_A value is not set to an integer; according to Figure 1, m_A can be obtained by Equation (5).

$$m_A = m + \frac{\delta_2}{\Delta\theta} + \frac{\delta_3}{\Delta\theta} = m + \frac{\delta_2}{\delta_1 + \delta_2} + \frac{\delta_3}{\delta_3 + \delta_4} \tag{5}$$

where $\delta_2 = \theta_i - T_i$ is the time difference between the i sampling point (θ_i, Y_i) and zero point $(T_i, 0)$, and $\delta_3 = T_{i+1} - \theta_{i+m}$ the time difference between the last sampling point (θ_{i+m}, Y_{i+m}) and the zero point $(T_{i+1}, 0)$. $\delta_3 = T_{i+1} - \theta_{i+m}$. Since we define the time difference of two adjacent sampling points, $\Delta\theta = T_s$, the total sampling points corresponding to one period (T) from $(T_i, 0)$ to $(T_{i+1}, 0)$, are equal to m_A. Therefore, the entire period of sine wave can be a fraction of the number of samples, which makes signal period calculation more accurate. Applying the definition of equilateral triangle in Figure 1, the following formula can be obtained

$$\frac{\delta_2}{\delta_1 + \delta_2} = \frac{|y_i|}{|y_i| + |y_{i-1}|} \tag{6}$$

$$\frac{\delta_3}{\delta_3 + \delta_4} = \frac{|y_m|}{|y_m| + |y_{m+1}|} \tag{7}$$

Equation (5) can be expressed as

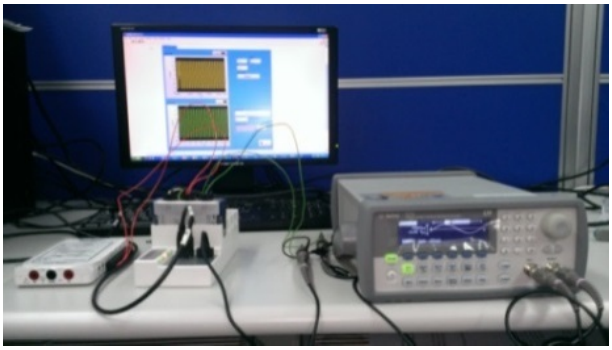
$$m_A = m + \frac{|y_i|}{|y_i| + |y_{i-1}|} + \frac{|y_m|}{|y_m| + |y_{m+1}|} \tag{8}$$

The accurate frequency of the evaluated signal can be estimated as

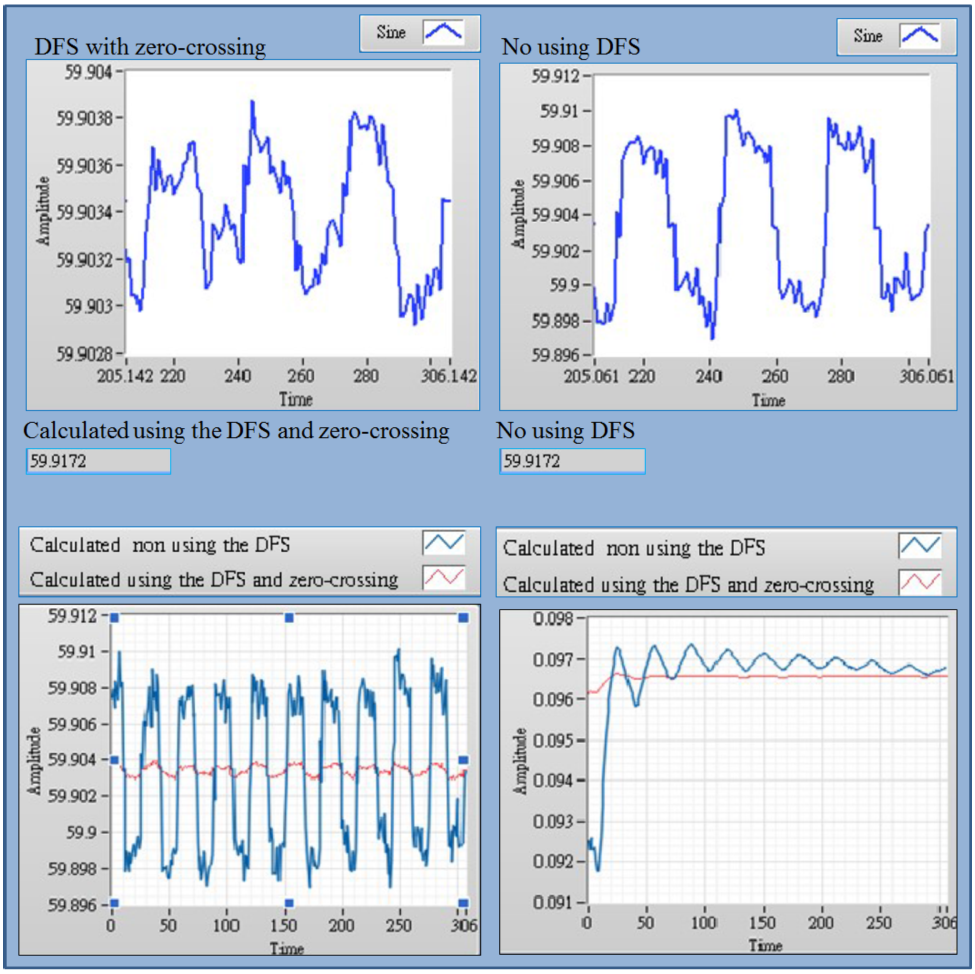
$$f = \frac{1}{m_A T_s} = f_a \frac{m}{m_A} \tag{9}$$

Even (θ_i, Y_i) and $(T_i, 0)$, (θ_{i+m}, Y_{i+m}) and $(T_{i+1}, 0)$ are exactly overlapping each other, Equations (8) and (9) can still be used for frequency detection.

3. **Hardware Setup and System Architecture.** Figure 2(a) is the hardware setup of this research where wiring setup for port AI0 of NI-9225 [10,11] is wired to the real power signal; port AI1 is the signal sent by NI-myDAQ, and port AI2 is signal sent by



(a)



(b)

FIGURE 2. (a) Experiment hardware setup, (b) comparison for Zero-crossing algorithm with and without DFS

Function-generator. The Function-generator can generate a specific signal to test the measure accuracy of NI-9225 and also can be correlated with the signal output from NI-myDAQ. Figure 2(b) shows the comparative diagrams for the power signal amplitude using Zero-crossing algorithm with and without Discrete Fourier Series (DFS).

The block diagram of the overall system is shown as Figure 3. There are one master node and six slave nodes to manage data flow. The name node and data nodes are corresponding to Apache HDFS (Hadoop Distributed File System) [12] architecture which is responsible for the data discovery and storage. The ResourceManager and NodeManager in YARN Layer [13] are devoted into computing jobs. It should be noted that the slaves include HDFS data nodes and YARN NodeManager, and the master is composed of HDFS name node and YARN ResourceManager.

4. **Data Processing Flow.** The power signals together with the weather data are collected from different sites as shown in Figure 4. After necessary ETL (Extraction, Transformation, and Loading) processing, the power signal data are stored into MongoDB [14] database. The filtering mechanism is responsible to pick up the special power event and then feed those into the analytic engine for further analysis such as clustering and classification. In this paper, we use analytic tools for the R Project packages [15].

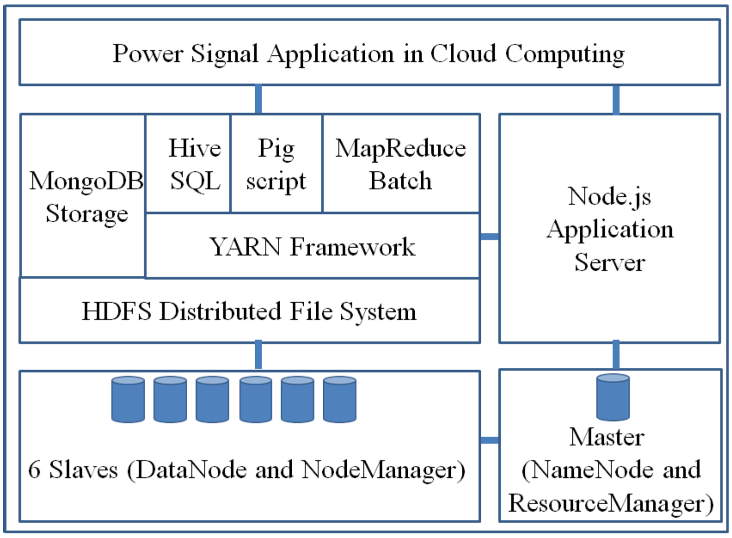


FIGURE 3. Overall system functional block diagram

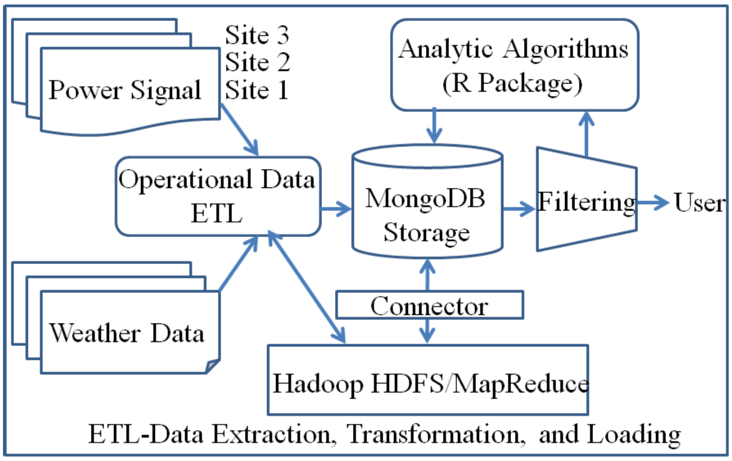


FIGURE 4. Collecting and processing power signal

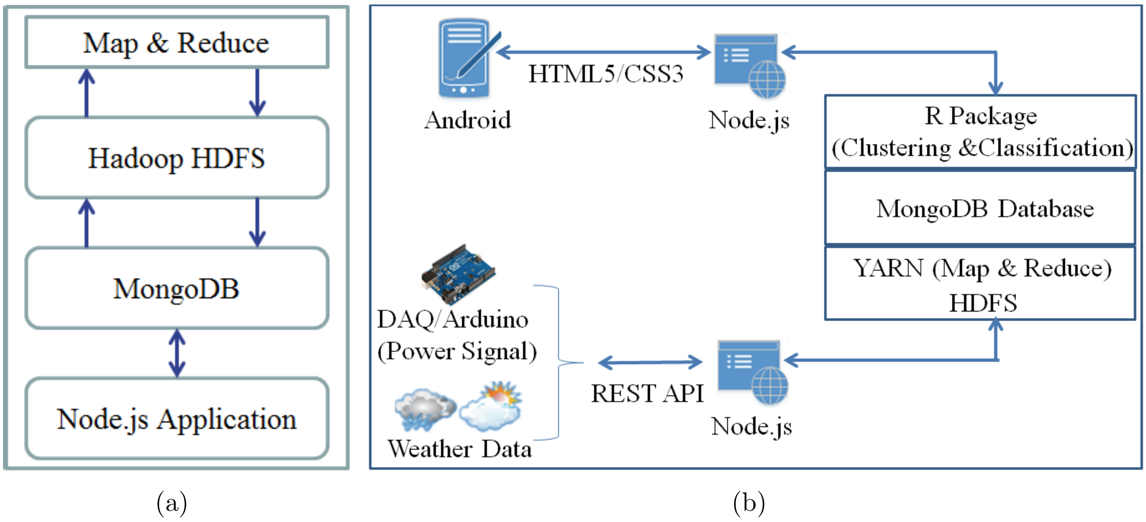


FIGURE 5. (a) Web access diagram, (b) detailed web access diagram

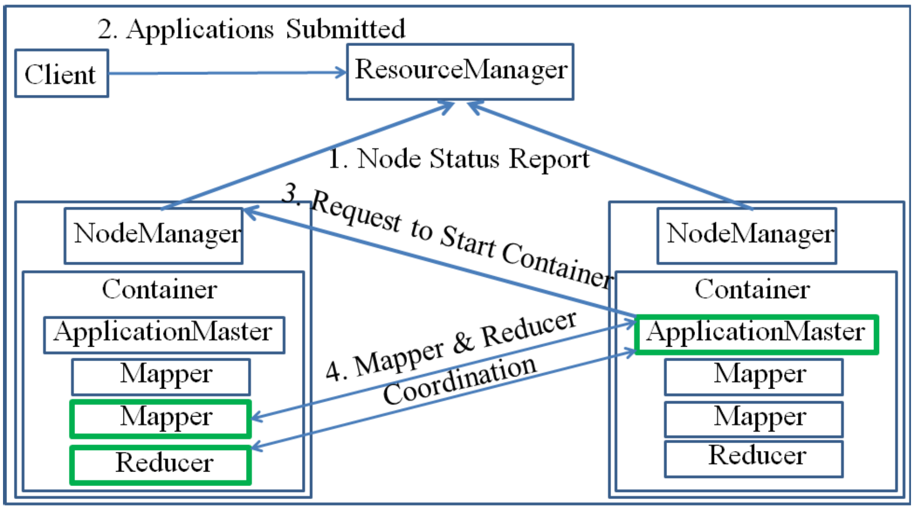


FIGURE 6. Data flow in YARN framework

In order to receive power signal from sensors and to present Web access to Android cell phone, we establish the application server using Node.js framework to process the power data as shown in Figures 5(a) and 5(b). The power signal data together with the weather information are sent into Apache HDFS to record the unusual power event. Apache HDFS is a distributed file system which accommodates the raw data to be computed by Mapper and Reducer tasks. By the way, after Mapper and Reducer computations, the frequency and the corresponding RMS value of the power signal data are stored into the MongoDB database for data mining.

YARN NodeManager in YARN framework as shown in Figure 6 is a per-node process which manages the resource assignment for each computing node. The resource request from different applications (also known as jobs) is submitted to YARN ResourceManager and then YARN NodeManager is requested to start a new Container which an YARN ApplicationManager is resided in. In other words, YARN ApplicationMaster will coordinate with the YARN NodeManagers to start and monitor the corresponding application containers for the requested resource. It is worth noting that the management work of Mapper and Reducer is located on YARN ApplicationMaster rather than the YARN ResourceManager. YARN ApplicationMaster which is deployed first for each application is responsible for coordinating the computation of the Mapper and Reducer.

5. **Conclusions.** The electric power measurement algorithms based on Zero-crossing algorithm are simple and fast. We first propose Fourier transformation algorithm to filter out the higher order harmonic and noise of the signal and then apply the Zero-crossing algorithm to calculate the fundamental frequency of the power signal accurately. Moreover, this article indicates the distributed implementation of power signal analysis system based on cloud computing architecture. For this cloud computing architecture, Apache Hadoop HDFS and YARN are used as file system and computing framework. In addition, the MongoDB acts as the database for the storage of the power signal raw data. Finally, the data mining tools in the R Project packages are used to find out the unusual power events. Such a cloud computing system using open source software packages is not only useful for power quality measurement but also suitable to be applied on many big data analysis issues.

Acknowledgement. This work was supported by the National Science Council of Taiwan under Grant NSC 102-2221-E-468-021-.

REFERENCES

[1] R. C. Dugan, M. F. McGranaghan and H. W. Beaty, *Electrical Power Systems Quality*, McGrawHill International Editions, 2000.
[2] D. O. Koval and C. Carter, Power quality characteristics of computer loads, *IEEE Trans. Industry Applications*, vol.33, no.3, pp.613-621, 1997.
[3] T. Q. Tran, L. E. Conrad and B. K. Stallman, Electric shock and elevated EMF levels due to triplen harmonic, *IEEE Trans. Power Delivery*, vol.11, no.2, pp.1041-1049, 1996.
[4] A. A. Girgis and W. L. Peterson, Adaptive estimation of power system frequency deviation and its rate of change for calculating sudden power system overloads, *IEEE Trans. Power Delivery*, vol.5, pp.585-594, 1990.
[5] I. Kamwa and R. Grondin, Fast adaptive scheme for tracking voltage phasor and local system frequency in power transmission and distribution systems, *IEEE Trans. Power Delivery*, vol.7, pp.789-795, 1992.
[6] IEEE recommended practice for monitoring electric power quality, *IEEE Std.1159-2009*, 2009.
[7] N. P. Yu, S. Shah, R. Johnson, R. Sherick, M. Hong and K. Loparo, Big data analytics in power distribution systems, *Proc. of IEEE PES Conference on Intelligent Smart Grid Technology*, Washington DC, 2015.
[8] Z. Huang, T. Gao, H. Zhang, X. Han, J. Cao, Z. Hu, S. Yao and Z. Zhu, Transient power quality assessment based on big data analysis, *China International Conference on Electricity Distribution*, Shenzhen, China, 2014.
[9] N. Liu and J. Zhou, The research and application of a big data storage model, *International Journal of Database Theory and Application*, vol.8, no.4, pp.319-330, 2015.
[10] *Dada Sheet of NI 9225*, NI CompactDAQ, https://sine.ni.com/nips/cds/view/p/lang/zht/nid/208795.
[11] *Dada Sheet of NI cDAQ-9174*, NI compactDAQ, USB Data Acquisition Systems, https://sine.ni.com/nips/cds/view/p/lang/zht/nid/207535.
[12] *Apache HDFS Website*, https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html.
[13] *Apache YARN*, http://hadoop.apache.org/docs/current/hadoop-yarn/hadoop-yarn-site/YARN.html.
[14] *MongoDB Website*, https://www.mongodb.org/.
[15] *R Project Website*, http://www.r-project.org/.