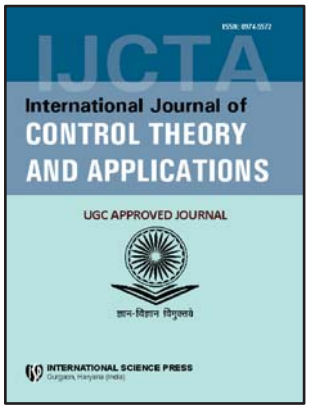


International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 10 • Number 30 • 2017

Smart-Frame for Smart Grid Data: A Secure Cloud Computing Framework for Big Data Analysis on Smart Grid Data

Abhishek Nigam[a] and Godwin[a]

[a]Dept. of Information Technology(Cloud Computing), SRM University, SRM Nagar, Potheri, Kattankulathur, Kancheepuran- 603203,Tamil Nadu,

E-mail: India abhisheknigam023@gmail.com, godwin.j@srmuniv.ac.in

Abstract: Smart Grid is an electronic grid which uses information and communication technology for gathering data related to the behaviour of supply and demand pattern of power and to act accordingly upon it in an automated fashion. While electricity demand and supply requires forecasting on a large, industrial and national level, this paper focuses on short-term (1 and 24 hour ahead) electricity demand forecasting for substations at the generation, distribution and transmission level. In this paper, we propose an efficient, secure and effective cloud based framework using cloudsim, Advance Encryption Standard and Hadoop Framework respectively. Here the data will get stored in cloud storage and will be encrypted using AES-256 encryption scheme. The data analytics techniques like Linear analysis and Predictive analysis will be performed on this data after decryption.

The existing model uses Cloud framework to store and manage Smart Grid data using clusters of cloud. This increases the cost as well as the complexity of managing cloud clusters, therefore it is not efficient. There are various existing security schemes like Identity based re-encryption scheme for securing the smart grid data but, is less efficient for processing huge amount of data. Since, AES-256 is strong and unbreakable till date thus, it is the better and efficient encryption scheme to be used.

Keywords: *Data Encryption, Identifying Parameters, Data Analytics, Linear Analysis, Predictive Analysis.*

1. INTRODUCTION

With the ever-increasing need for generation of electricity worldwide, electricity has greater impact towards the industrial development of a country. The Smart Grid considers and analyzes the historic data of electricity to predict whether the supply will meet the demand or not. It also takes alternative decision for meeting the demand. The growth in database will thus increase and will be available as the historic data which is helpful for future analysis .Here, the data is being collected through a random-generator which is being treated as historic data that gets stored in the cloud created using cloudsim. Based on the pre-determined parameters the data needed for prediction of future demand-supply pattern by the Smart Grid is retrieved.

The programming language used for accessing and analyzing data is java. Entire data analysis of data in cloud will be done using java as it is more effective and platform independent. The aim of this paper is to propose a secure cloud-based framework which automates the system to maintain the balance between the demand and supply.

2. THE BIG DATA ANALYTICS

It is a field which deals with data which are high in volume, velocity and variety. The usage of big data analytics becomes mandatory when the data to be analyzed is larger in volume, or the velocity of the incoming data is higher or the data is variable in nature. Big data analytics comprises of distributed file system framework such as Hadoop. which are open source in nature from Apache Foundation.

2.1. The Hadoop Distributed File System (HDFS)

Has a master-slave architecture, designed to run on commodity hardware. A HDFS cluster consists of a single Name Node and a master server that manages the file system namespace and regulates access to files by client. A Data node usually one per node is present to manage the storage in the cluster. In HDFS a file is split into one or more blocks which are in turn stored in Data Nodes. The Name Node handles the namespace operations such as opening, closing and renaming directories and also maps the blocks to Data Nodes. HDFS is highly fault-tolerant and is designed to be deployed on low-cost hardware.

3. SOFTWARE REQUIREMENT

Front End: Java swing

Back End: HDFS

Software: Java JDK 1.8, Hadoop 1.2.1, Cloudsim 3.0.3

OS : Ubuntu 12.04 LTS 64 bit

IDE : Netbeans 8.0.1

3.1. Modules

1. Cloud computing environment
2. Encryption of smart grid data
3. Storing smart grid data in HDFS
4. Retrieving smart grid data from HDFS and Decryption
5. Predictive Analysis over decrypted data

3.1.1. Cloud computing environment

It is a generalized simulated framework. Cloudsim is used to simulate the cloud-based infrastructures (Data centre, Host, Vm) and application service. To model the resource data centre, variation of CPU processors, storages, bandwidths, operation systems, virtualization standards and machine locations are considered during the resources configuration. To model cloud consumers, application tasks are created in terms of cloudlets, including all information related to task execution management details such as tasks processing requirements, expressed in MIPS, disk I/O operations and the size of input files.

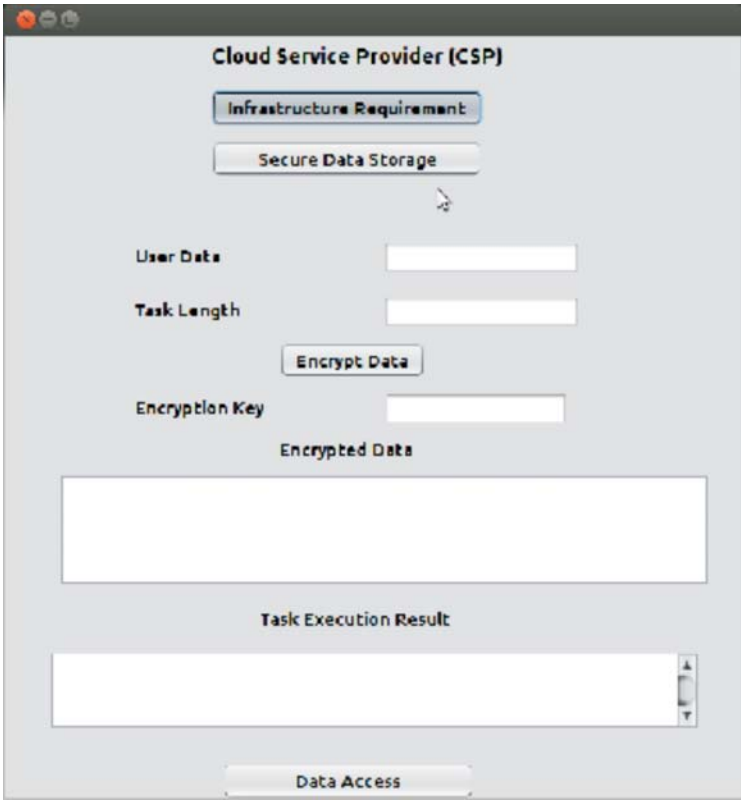


Figure 1: Cloud Environment for User Requirement

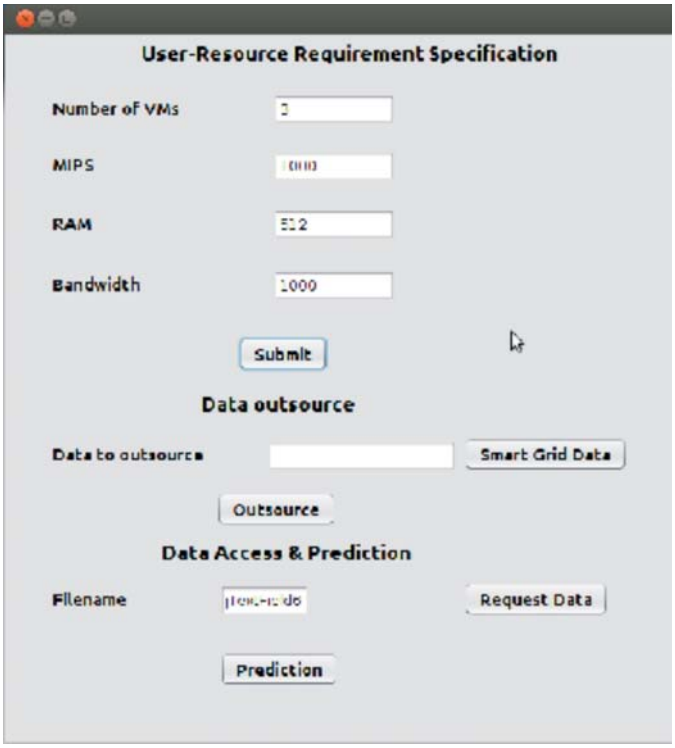


Figure 2: Specifications for cloud resources

3.1.2 Storing smart grid data in HDFS

Encrypted Smart-grid data is stored in HDFS. Storage process is shown as a cloudlet execution in cloudsim. Host object from cloudsim package is assumed to possess the HDFS in it.

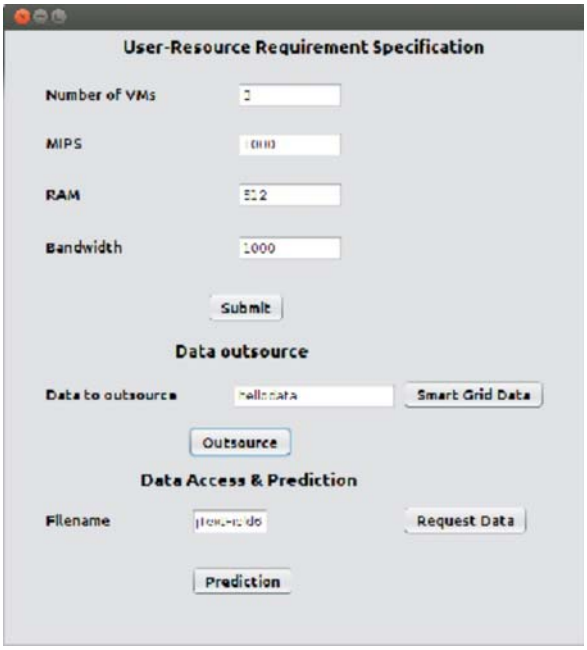


Figure 3: Selecting input file and sending it for analysis

3.1.3. Encryption of smart grid data

Smart grid data is randomly generated and it is encrypted using AES algorithm. File size of encrypted data is assigned as a file size parameter of cloudlet object.

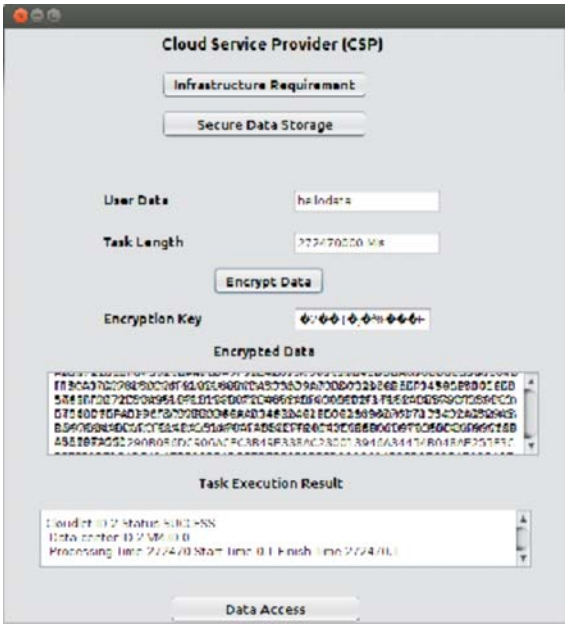
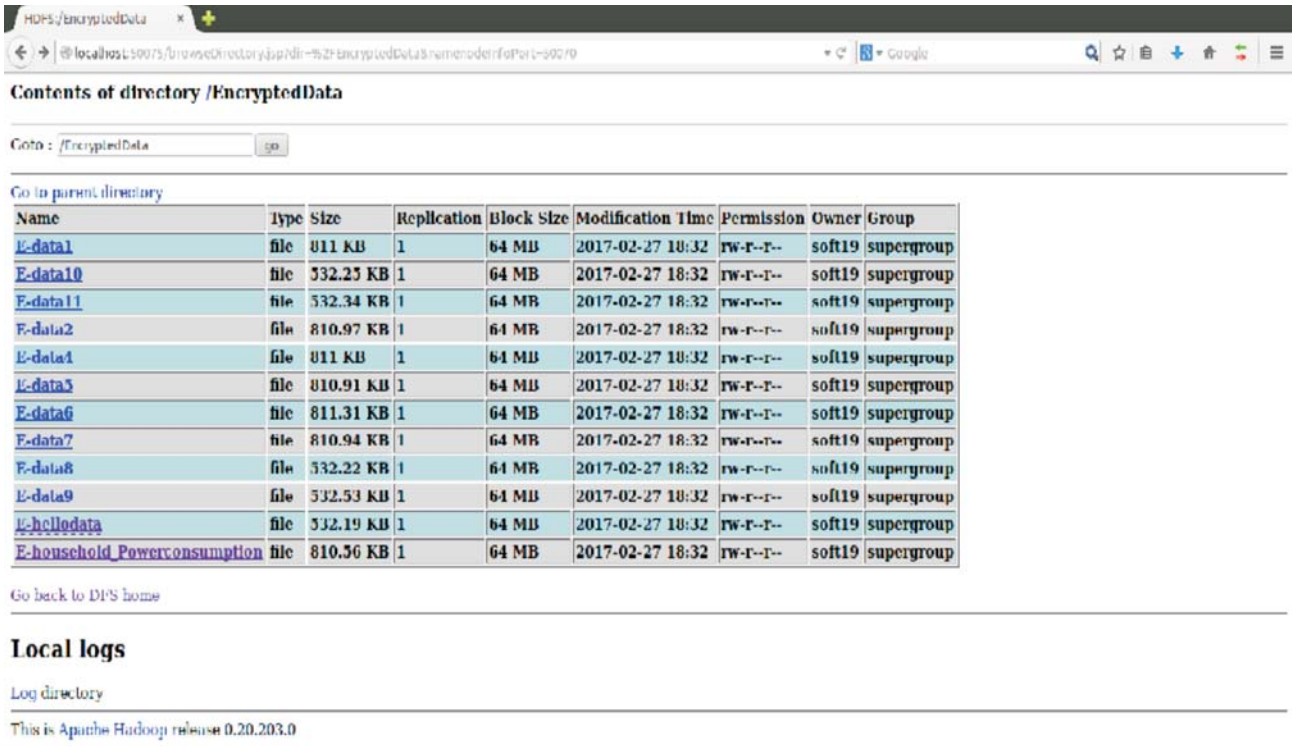


Figure 4: Data encryption



Contents of directory /EncryptedData

Goto : /EncryptedData go

Go to parent directory

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
E-data1	file	811 KB	1	64 MB	2017-02-27 18:32	rw-r--r--	soft19	supergroup
E-data10	file	532.25 KB	1	64 MB	2017-02-27 18:32	rw-r--r--	soft19	supergroup
E-data11	file	532.34 KB	1	64 MB	2017-02-27 18:32	rw-r--r--	soft19	supergroup
E-data2	file	810.97 KB	1	64 MB	2017-02-27 18:32	rw-r--r--	soft19	supergroup
E-data4	file	811 KB	1	64 MB	2017-02-27 18:32	rw-r--r--	soft19	supergroup
E-data5	file	810.91 KB	1	64 MB	2017-02-27 18:32	rw-r--r--	soft19	supergroup
E-data6	file	811.31 KB	1	64 MB	2017-02-27 18:32	rw-r--r--	soft19	supergroup
E-data7	file	810.94 KB	1	64 MB	2017-02-27 18:32	rw-r--r--	soft19	supergroup
E-data8	file	532.22 KB	1	64 MB	2017-02-27 18:32	rw-r--r--	soft19	supergroup
E-data9	file	532.53 KB	1	64 MB	2017-02-27 18:32	rw-r--r--	soft19	supergroup
E-hellodata	file	532.19 KB	1	64 MB	2017-02-27 18:32	rw-r--r--	soft19	supergroup
E-household_Powerconsumption	file	810.56 KB	1	64 MB	2017-02-27 18:32	rw-r--r--	soft19	supergroup

Go back to DFS home

Local logs

Log directory

This is Apache Hadoop release 0.20.203.0

Figure 5: Encrypted file Storage in HDFS

3.1.4. Retrieving smart grid data from HDFS and Decryption

Encrypted Smart-grid data is retrieved from HDFS and it is decrypted using AES algorithm. File size of encrypted data is assigned as a file size parameter of cloudlet object. Retrieval process is shown as a cloudlet execution in cloudsim. Host object from cloudsim package is assumed to possess the HDFS in it from which data is retrieved.

3.1.5. Predictive Analysis over decrypted data

Decrypted data is processed for predictive analysis [1] (See Fig 6)

3.1.5.1. Algorithm for Predictive Analysis

1. Construct a Regression Model to fit a straight lline to the data set.
2. The data is provided in the form of *x* and *y* points.
3. Get the coefficient of the fitted straight line,

   Covariance = (first_element_of_data[*i*] – *x*mean)*(second_element_of_data[*i*] –*y* mean)
4. Get the variance of the data set,

   Get the mean of the data set (sum the difference between the data element and the mean squared)

   sum Of Squared Deviations = sum Of Squared Deviation + Math.pow(data[*i*] – mean,2)

   Divide the sum by the length of the data set -1 to get our result as,

   sum Of Squared Deviations/(data.length -1);

5. Get the value of the gradient using the formula
   $b = \text{cov}[X,Y]/ \text{var}[X]$
6. Get the value of the *y*-intercept using the formula
   $a = y' + b*x'$
7. Evaluate the computed model at a certain point.
8. Return the value of the fitted straight line at the point *x*
9. Return $a + b * x$;

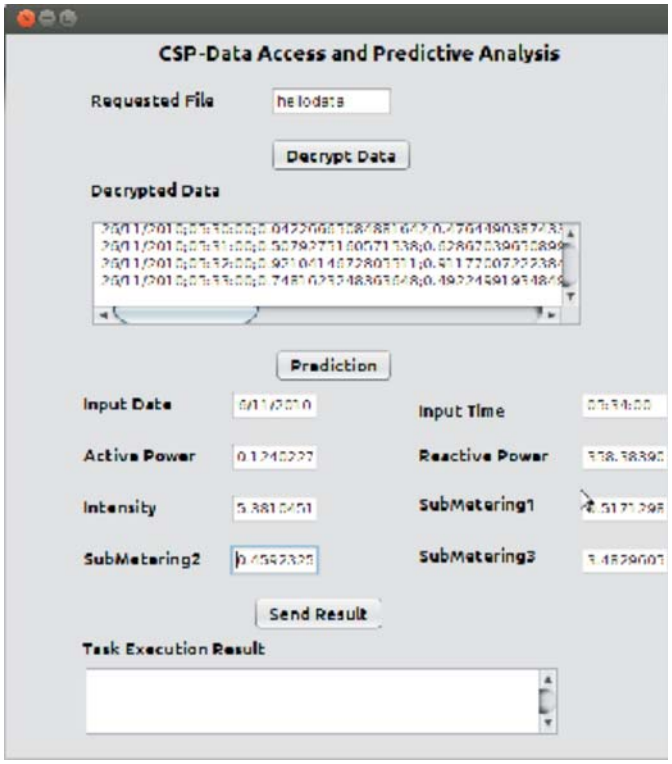


Figure 6: Decryption on data once fetched from HDFS and Prediction Result

4. BACKGROUND

Smart Grid is a technological innovation that improves efficiency, reliability, economics and sustainability of electrical services. It plays an important role in establishing a dominant energy infrastructure. The most important challenge for smart grid is to manage and process huge amount of data it receives from the electrical devices. Cloud computing is a good candidate for addressing these challenges since it is energy efficient, cost efficient, posses agility, scalability and flexibility. In the existing paper, the proposal is of a secure cloud computing based framework for data analysis in smart grid.

4.1. In our paper

We present a secure and efficient way of securing data in the cloud and performing data analytics on it. Here, data is as it enters the cloud store is first encrypted using AES-256 algorithm and is then stored in the encrypted format. Now, the parameters which have a visible effect on the demand and supply pattern are identified, on the basis of which the data is fetched and decrypted so that it is available for Hadoop framework. Data analytics is then performed on this data by means of linear and predictive analysis. The results are then stored in separate file for future references.

REFERENCES

[1] D. Frincle, M.Hadley, H. Khurana,and N. Lu "Smart-grid security issues" IEEE Security Privacy, Jan-Feb **2010.**

[2] X.Fang, S. Misra, G.Xue and D. Yang "Managing smart-grid information in cloud: Opportunities, model and applications", IEEE Network, Jul-Aug **2012.**

[3] A. Jwayemi, et.al., "Knowing When to Act: An optimal Stopping Method for Smart Grid Demand Response", IEEE Network, Sep-Oct **2011.**

[4] CloudSim installation Tutorial,[online] Available: http://wahnilla.wordpress.com/category/cloudsim-tutorials/, http://cloudsim-setuo.blogspot.in/2013/04/install-and-rum-cloudsim-321-using.html?m=1/.

[5] CloudSim Tutorial, [online] Available: http://www.superwits.com/.

[6] Apache Foundation Hadoop HDFS Architecture, Available: http://hadoop.apache.org/docs/stable/hdfs_design.htm.

[7] Apache Cassandra Tutorial,[online] Available: http://cassandra.apache.org/, http://www.tutorialspoint.com/cassandra/cassandra_introduction.htm.

[8] Apache Hive Tutorial, [online] Available: http://hive.apache.org/.

[9] Sharifi, Leila, Felix Freitag and Luis Veiga "Combining Smart Grid with community clouds: Next Generation integrated service platform". **2014** IEEE International Conference on Smart Grid Communications (SmartGridComm), **2014.**

[10] lawreview.vermontlaw.edu.

[11] Razia Sulthana, A., ana Ramasamy Subburaj "An improvised Ontology based K-Means Clustering Approach for Classification of Customer Review", Indian Journal of Science and Technology, **2016.**

[12] Privacy Preserving Predictive Analytics with Smart Meters, **2016** IEEE International Congress on Big Data.

[13] Fei Teng. "A New Game Theoretical Resource Allocation Algorithm for Cloud Computing", Lecture notes in Computer Science, **2010.**