SETSCI Conference Proceedings, 4 (1): 367-373, 2019

ISAS 2019
3rd International Symposium on Innovative Approaches in Scientific Studies
Section: Engineering and Natural Sciences
April 19 – 21, 2019, Ankara, Turkey

Efficient Data Storage in Edge Cloud Computing For Real Time System

Alhakam Qassab[1*], Dr. Shafqat Ur Rehman[2+]

[1] Computer Engineering /Research Institute, Ankara Yildirim Beyazit University, Ankara, Turkey
[2] Computer Engineering /Research Institute, Ankara Yildirim Beyazit University, Ankara, Turkey
[*]Corresponding author: hakamsameer1988@gmail.com
[+]Speaker: hakamsameer1988@gmail.com
Paper Type: Full Paper

Abstract – At the last two decades the information world and associated information technologies have been progressed. This progress can be implemented from the concept of the internet to the internet of things. The Internet of things (IoT) refers to billions of electronic devices around the world that are lınk up to each other's by using wireless communication networks techniques, such as Bluetooth, ZigBee, Wi-Fi, 2G, 3G, 4G mobile protocol and for example about (IoT) Smart infrastructure, Smart city, Smart Mobility, Smart technology, etc...

By using like these smart systems will generate a huge of data sensing that captures from the physical world by sensors and will need to store and process in the centralized process (cloud computing) at less time in order to get faster feedback data response. Today the data analytics, which utilized on edge cloud computing has more and more important in order to get near real-time decisions.

The Edge cloud computing is a new paradigm from a cloud computing environment which provides support the data distributed processing that store and process the data sensing in the node to stay close from data produced resources instead of sending it to centralized processing (cloud computing) to minimize the data latency and investigate a near real-time response.

Because of limitation, storage on edge cloud computing, I am facing challenges to balance between the quality of data and the quantity of data stored on edge computing for taking near real-time decisions. In this thesis, I use three architectural layers for efficient data storage and management of edge cloud computing that include an adaptive algorithm that dynamically finds a trade-off between providing high prediction accuracy necessary to improve a real-time decision and decreasing the amount of data stored in limited space storage. The aim of this thesis focuses on time series data to get a near real-time decision.

.

Keywords – Internet of Thing, Cloud Computing, Edge Cloud Computing, Intelligent home system, Adaptive algorithm.

I. Introduction

In the last decade with extent using the internet, technology has been more and more intelligent and has been showing the internet of things at first time, according to Kevin Ashton, who is a first-person likely coined the term "Internet of things" later MIT's Auto-ID Center in 1999 [1].

Internet of things refers to a hundred or thousands of physical devices that communicate with each other by using wireless communication technologies and can conduct it remotely in order to share date sense. The aim of the internet of things is to enhance the quality of services (QoS) through support the applications which used in the daily life, for example, of the internet of things smart city, smart communications equipment and smart health equipment like these smart things will generate a huge data that need to store and process.

Today the data analytics, which utilized on edge cloud computing has more and more important in order to get a near real-time decision. The edge cloud computing is the last version from the cloud computing that supports the distributed process which allows for analysis the data sensing close from data resources, By putting data analytics close to the source of data, Edge architectures can decrease the amount of data traversing the network, thus minimizing latency and overall Costs. In Addition to that, edge computing allows getting fast and accurate responses.

Because of limitation, storage in edge cloud computing we cannot store all of the data which generate from internet of things to implement real-time analytics, we need a lot of historical data to select the high accurate predictions, thus improve storage of real-time analytics has become majority issue on edge cloud computing.

To improve the data management and analytics on edge computing we will use three architectural layers which include an adaptive algorithm that will find a trade-off between decreasing the necessary data storage and accuracy of forecasts

The rest of the paper is organized as follows. Internet of things (IoT), Cloud Computing, Edge Computing, Intelligent home system, adaptive algorithm.

II. Internet of Things

In last few years and with extend using the internet and development the day life internet of things term has been known, Internet of things is a represent emerging technology that involves of billions physical devices or nodes around the world that connect to each other by using wireless communication network techniques. The main aim of the internet of things (IoT) is to improve the quality of service (QoS) through sharing exchange and analyses a massive data between the smart software or a smart hardware over the internet at less time, For example of these technologies (Smart city devices, Smart control system, Smart mobility Smart technology etc…)[2][3].

Fig. 1 Example of Internet of things

III. CLOUD COMPUTING

Recently new paradigm has emerged in order to hosting and delivering the services over the internet and sharing pool computing resources such as (networking, storage, application services). The cloud computing allows us to create, configure and customize the applications over the internet.
The cloud computing (CC) provides flexibility and efficiency of using and accessible to resource computing.
The major goal of cloud computing (CC) is enabling the users from using the computing resource from anywhere as you wants without to need from download the platform on a desktop **[4]**. We can geagorize the cloud computing services into three kinds. As is illustrated in figure 2.

- Infrastructure as a Service (IaaS).
- Platform as a Service (PaaS).
- Software as a Service (SaaS).

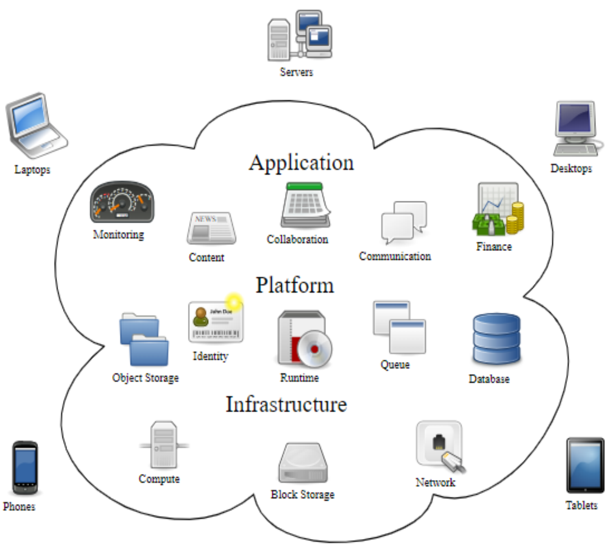


Fig. 2 Example of the Cloud Computing

IV. EDGE CLOUD COMPUTING

Edge cloud computing is a new emerging paradigm architectural which helped from resolving some cloud computing problems such as high cost and latency by pushing computing services from the centralized process (cloud computing) into the geographical process and make it close from data sources. The aim from the edge cloud computing is to get faster feedback data response in order to make a near real-time decision.

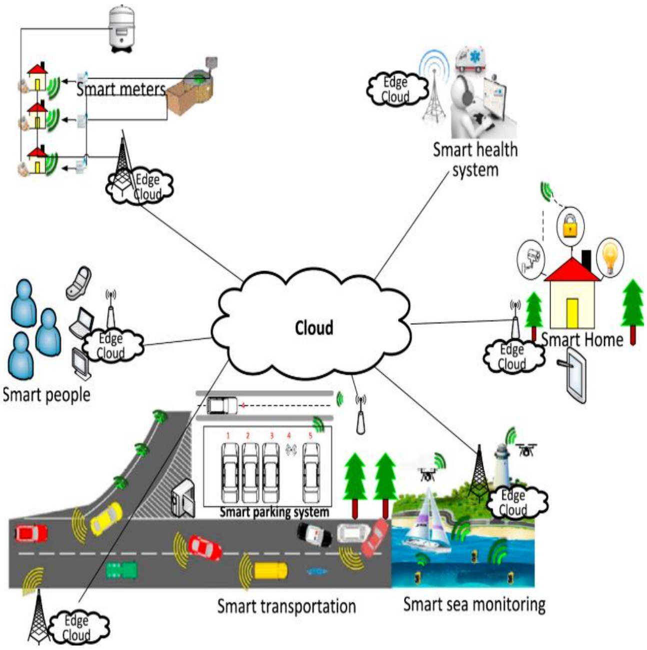


Fig. 3 Example of Edge Cloud Computing

V. EDGE COMPUTING FEATURES

- Reduce the latency.
- Support Mobility.
- Geographical processes.

VI. INTELLEGENCE HOME SYSTEM

In this paper we used MATLAB programming languages for implementing the algorithm. For evaluation of the results in proposed approach we used the amount of data and the accuracy of the forecasts which has to save in the Edge node. For dataset we used the UMass Trace Repository [6]. This dataset has traces coming from the Smart* project [7] for resolve of scheming sustainable homes. These traces represent real data and has potential information for IoT actuators which varieties these datasets appropriate samples for proposed experiments. In this paper the three datasets targeting are used. Also for showing the applicability of proposed algorithm different characteristics of datasets are used.

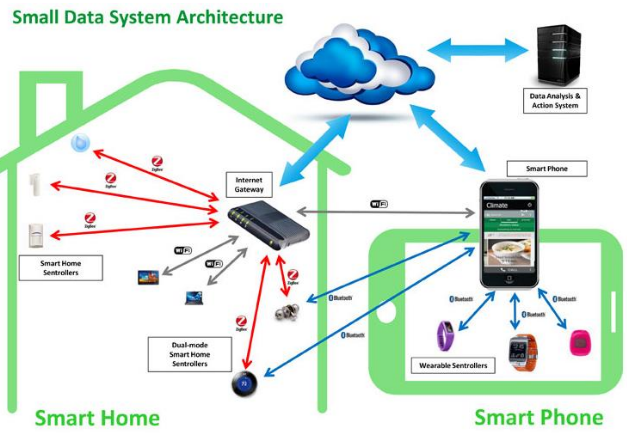


Fig. 4 Example of smart home system

Edge Computing is a new computing model in this reverence that is foreseen to be a part of this change in addition to the Internet of Things (IoT), Machine Learning (ML) and Big Data Analytics. The target of the Edge Computing is move computing from distant cloud environments to the edges of the networks. Recently in modern cities have used the centralized process (cloud computing) for helping to make a safe decision, but a massive of data that generated from smart IoT home also will face a lot of challenges, such as analysis, searching, storing, sharing, transfer, updating, querying. According to Intel flag records a home requires to analysis a huge sensor data approximately 1 GB/S for making a safe decision. We can classify the big data processing that using in the intelligent transportation system into two kinds.

1- Centralized big data processing.
2- Distributed big data processing.

The centralized big data processes will be difficult to provide feedback response quickly for making a near real-time decision, from another side the big data processing will utilize the power of cloud computing, as a result of that the chance to prevent damage from a dangerous event mostly will be missing while waiting to receive a safe decision from the controller in centralized processing (cloud computing).

That is why to improve safety on a home we must to convert the big data processing into geographically distributed processing (edge cloud computing). By using the edge cloud computing in the smart homes will reduce the data latency and will lead to get feedback data quickly, as a result of this it will improve a safety in home through enabling the controller from making faster right decision.

VII. ADAPTIVE ALGORITHM DESIGN

We classified the adaptive algorithm design into six steps

1. Learning phase
2. Validation of the specification list
3. Multiple forecast iteration on available dataset
4. Detection of stable accuracy clusters
5. Data management action
6. Validation of available dataset

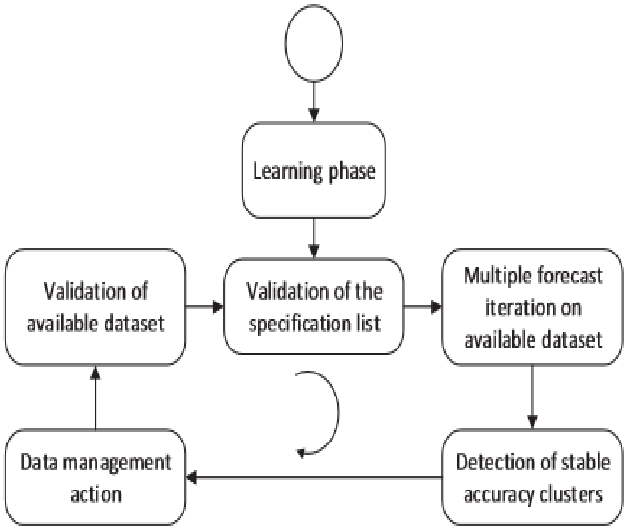


Fig. 2 Example of the Cloud Computing

1. Learning phase step:

The aim from the learning step is to derive the information from data set which used at first time in algorithm execution, in addition of that this step includes some actions such as time series pattern recognition that useful to determine the most appropriate accuracy forecast method

2. Validation of the specification list:

Validation of the specification list is the second step in adaptive algorithm, in this step will put the specification list such as forecast accuracy there should which specified during algorithm preparation and according to it will derive the best data, but any change at specification list will influence to all edge cloud computing storage management, so we need to check the list specification during the execution each time during the execution.

3. Multiple forecast iteration on available dataset:

During this step will specify one of the forecast methods like (ETS or ARIMA) with specified in specification list. The data will fix or equal to the number of points that specified for a client in the specified list as they are used only for accuracy evaluation purposes.

4. Detection of stable accuracy clusters:

The aim from detection of stable accuracy clusters is to find the stable clusters of accuracy values within the vector returned by the multiple forecast iteration phase.

5. Data management action:

İn the data management action will check either releases storage or quires for more data to mediator component, so there are three cases for this purpose.

- We can reach stable cluster with desired accuracy with fewer amount of data.
- Forecast accuracy of stable cluster is higher with the increased amount of training data, for example forecast based on all available data from the storage, than mediator component can send require to the cloud data store for more data.

- None of resulting accuracy of stable clusters meets the specified accuracy threshold by the client, In this case data management action will select the one with highest forecast accuracy but obtained with less data points.

6. Validation of available datasets:

Algorithm checks storage for new data that are collected while the data management action is completed. In the next cycle both data received from data preparation and from the mediator component are included. The adaptive algorithm is continually repeated and consequently, it tries to keep the lower quantity of data on the Edge, while trying to keep a level of accuracy that is equal to the forecast accuracy threshold value in the specification list.

VIII. ARCHITECTURE FOR EDGESTORAGE MANAGEMENT

We represent the proposed three-layer architecture model for Edge storage management on Edge nodes. The goal of the presented Edge architecture is to devise a selfmanagement process to automatically manage limited Edge data storage without intervention of third party

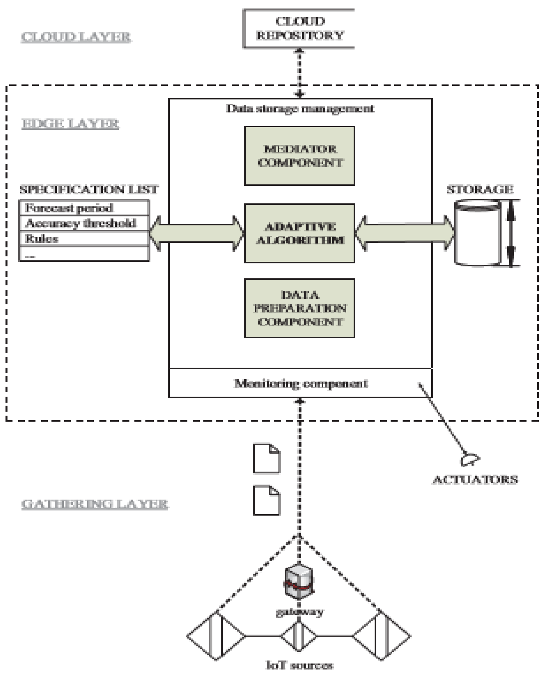


Fig 6 Example Architecture for Edge storage management

a) Cloud layer:
Represents a data repository which stores all historical data collected from monitored systems. This layer performs big data analytics and delivers useful information based on entire datasets.

b) Edge layer:
Manages data storage process based on our proposed adaptive algorithm. Further, this layer performs local analytics and extracts actionable information from available data. The edge layer consists of several components:

- Monitoring component:

Collects the data,monitors average amount of incoming data, monitors storage space, sends information on storage capacity and control commands to IoT actuators;

- Data preparation component:

Receives data from monitoring component and performs data preparation operations on these data (e.g., filling missing data, removing unnecessary data, normalizing data). After this step, the data are sent either to Edge data storage or to the Cloud layer via mediator component,

- Adaptive algorithm:

Provides the core of the data storage management. It receives data from the storage, checks the specification list, and implements design principles provided

- Specification list:

Consists of user-defined information for monitoring data and for adaptive algorithm such as: forecast period specified by client for monitoring process or by the edge system.

- Storage is responsible for:
  (1) Storing data received by data preparation component,
  (2) Sending data to the adaptive algorithm,
  (3) Receiving data retrieved by mediator component and
  (4) Storing adaptive algorithm results

- Mediator component:

At the Mediator component will requests from the cloud repository if the Mediator component needed range of data after that forwards them to the storage unit.

c) Gathering layer:
At the Gathering layer will Transmits (IoT) measurements to an Edge node, either indirectly by the gateway (in case of devices that are not capable of sending data through IP) or directly (IP-enabled devices). In order to reduce communication costs and latency in distributed sensor networks [10]

IX. SIMULATION RESULTS

For simulation result we used three data set from ***. For the first data set the result is shown in figure 7 and figure 8.

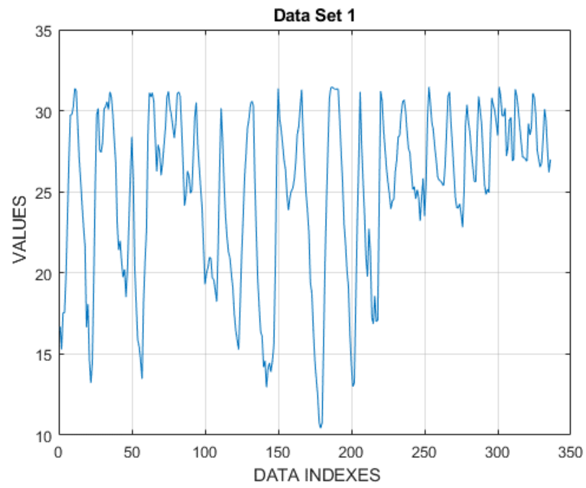


Fig. 7. Result of values vs. data indexes

The forecast accuracy is shown in figure 8.

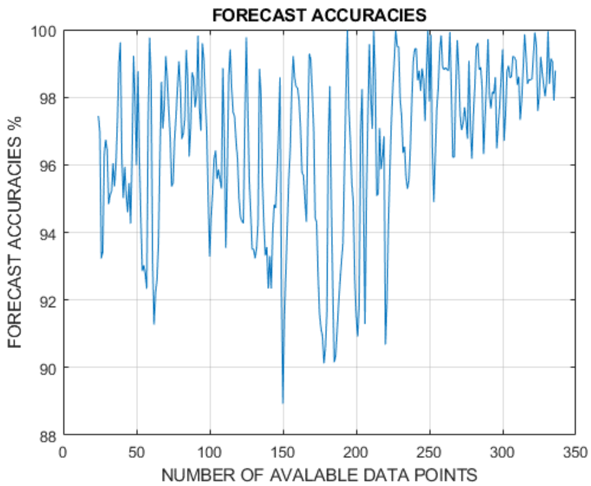
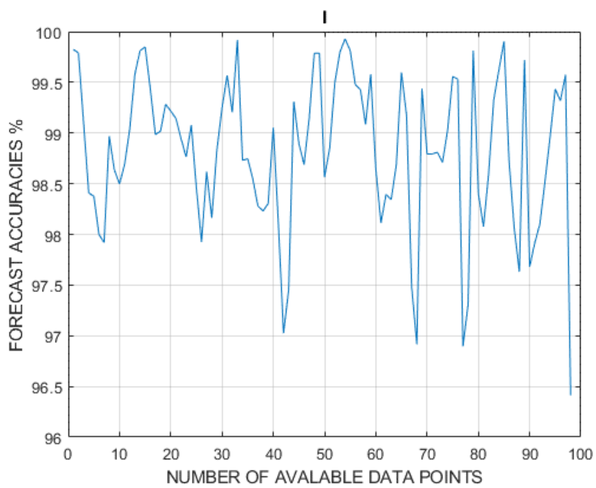
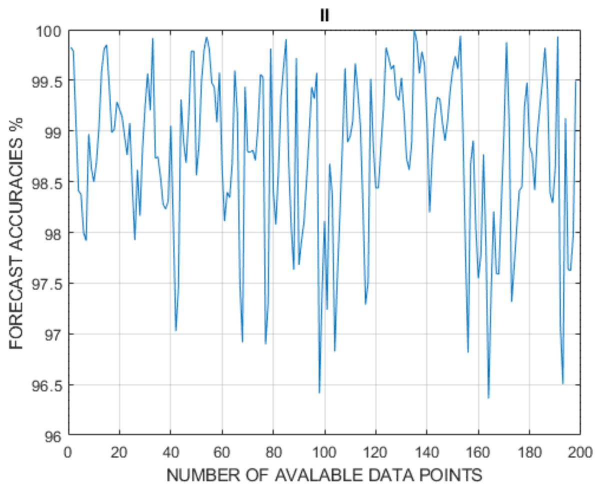


Fig. 8 Forecast accuracy vs. of number of available data points

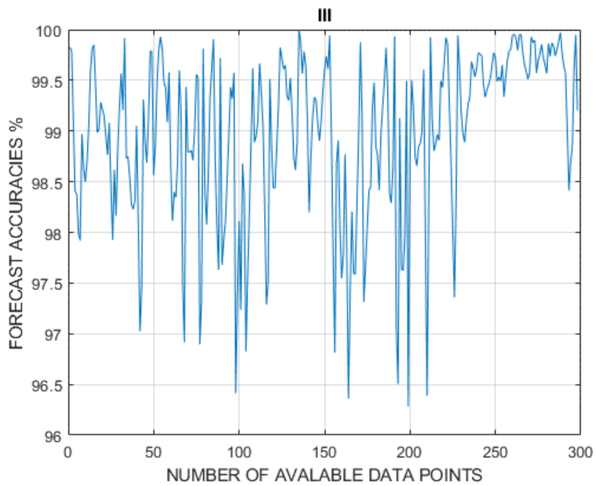
The Forecast accuracy (%) vs. number of available data points for 6 scenario in second data set is shown in figure 9.



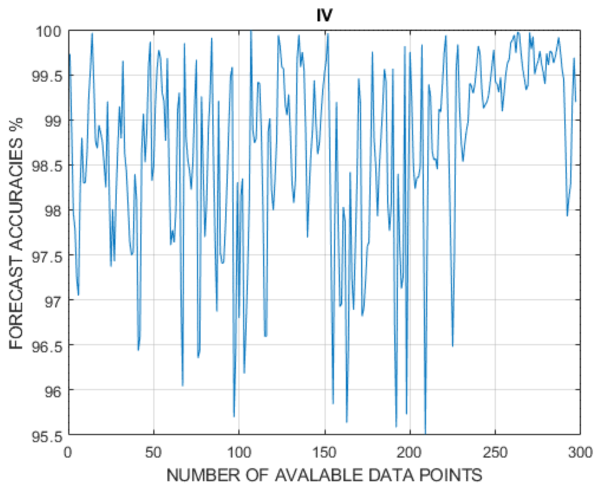
(a)



(b)



(c)



(d)

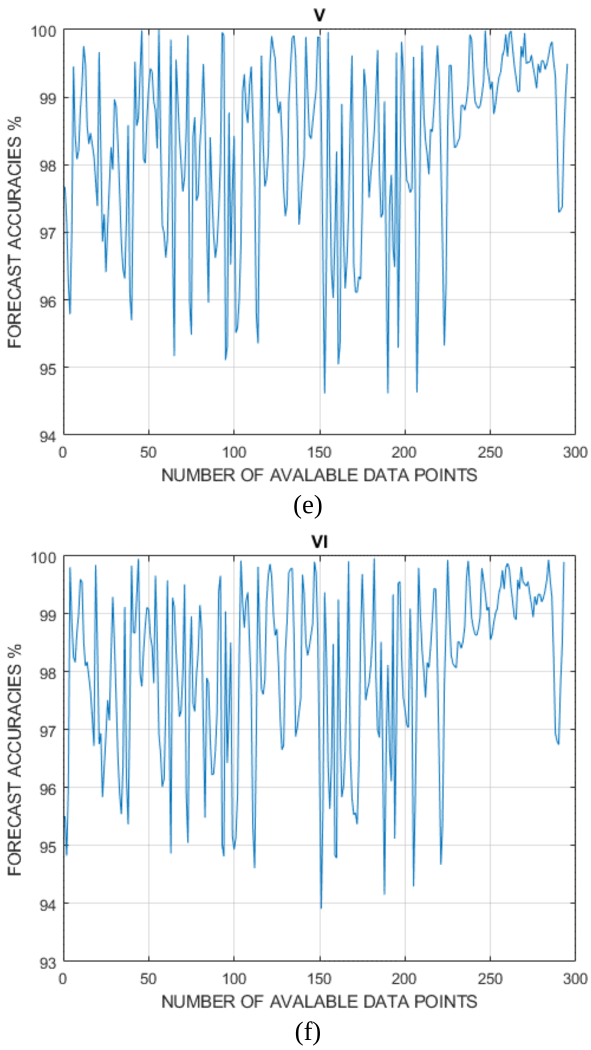


Fig. 9. Forecast accuracy (%) vs. number of available data points

For the third dataset the results shown in figure 10.

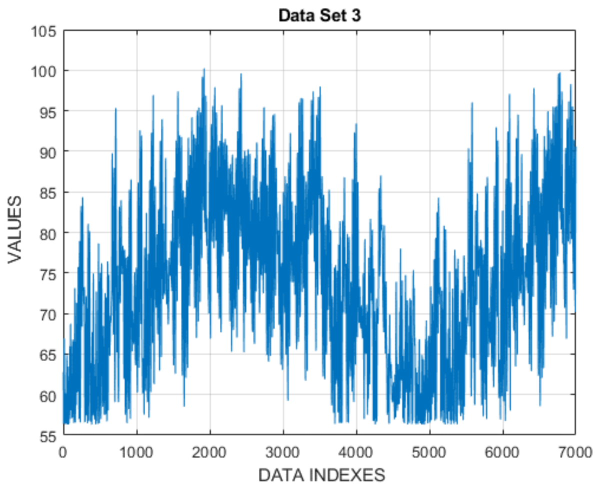


Fig. 10 Result of values vs. data indexes.

We focus on time series data, typical in the context of sensor-based monitoring in IoT environments. The results for 600 points dataset is shown in table 1.

Table 1. Results for 600 points dataset

Cycle	data	Accuracy (%)	data	Accuracy (%)	Available data	Accuracy (%)
I	100	98.8202	100	97.7242	100	97.0792
II	200	98.757	134	97.7242	134	97.6863
III	300	98.9597	219	96.7981	142	96.6915
IV	300	98.6447	214	96.7981	122	97.0792
V	300	98.2495	193	97.0525	136	97.6863
VI	300	98.0182	162	97.0608	148	96.9673
Average		98.5749	170	97.0783	148	97.6863

The stable clusters of forecast accuracy is shown in table 2.

Table 2. Stable clusters of forecast accuracy

Range of cluster		
Mean Value	**Index (Start)**	**Index (end)**
95.8979	278	281
99.0343	282	285
96.0204	286	289
99.4454	291	294
96.4416	296	300
96.1812	301	305
96.2567	312	315
99.6151	315	320
96.7847	323	326
99.7812	327	331

X. CONCLUSION

Our simulation results show that our adaptive algorithm can reduce the amount of data by an average 99.7812% for index start at 327 and 95.8979% for index start at 278 and end index at 281 for each cycle for the three datasets respectively, while satisfying demands for forecast accuracy and thereby showing potential for saving limited storage space.

REFERENCES

[1] https://www.dataversity.net/brief-history-internet-things/#J.

[2] Ashton, K. (22 June 2009). "That 'Internet of Things' Thing". Retrieved 9 May 2017.

[3] Y. Sun, H. Song, A. J. Jara, and R. Bie, "Internet of things and big data analytics for smart and connected communities," IEEE Access, vol. 4, pp. 766–773, 2016 (2002) The IEEE website. [Online]. Available: http://www.ieee.org/

[4] (Mathai, 2012) Madhavi, M. (2012). An Approach For Intrusion Detection System In Cloud Computing, 3 (5), 5219–5222.

[5] -Archive/macros/latex/contrib/supported/IEEEtran/

[6] "Umass trace repository," http://traces.cs.umass.edu/index.php/Main/Traces/, 2017, [Online; accessed 11-January-2017].

[7] S. Barker, A. Mishra, D. Irwin, E. Cecchet, P. Shenoy, and J. Albrecht, "Smart*: An open data set and tools for enabling research in sustainable homes," *SustKDD, August*, vol. 111, p. 112, 2012.

[8] A. Karnik, "Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP," M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.

[9] J. Lin, S. Williamson, K. Borne, and D. De Barr, Pattern Recognition in Time Series, Advances in Machine Learning and Data Mining for Astronomy. Chapman and Hall, 2012.

[10] J. L. Wong, R. Jafari, and M. Potkonjak, "Gateway placement for latency and energy efficient data aggregation [wireless SEnsornet works],"in29th Annual IEEE International Conference on Local Computer Networks, 2004, pp. 490–497.