**DR. B R AMBEDKAR NATIONAL INSTITUTE OF TECHNOLOGY JALANDHAR**

****

**Cloud Project Report**

**(Group No. 5)**

**Aug-Dec 2020**

|  |  |
| --- | --- |
| **SUBMITTED TO:** | **SUBMITTED BY:** |
| Dr. Mohit Kumar | Vikas Goyal (17103091) |
| CSE Department | Sudhanshu Gupta (17103084) |
|  | Abhinendra Gaurav (17103002) |
|  | Ankit Goyal (17103011) |

**Machine Learning Approach to Predict Workload at Fog Nodes**

# Abstract:

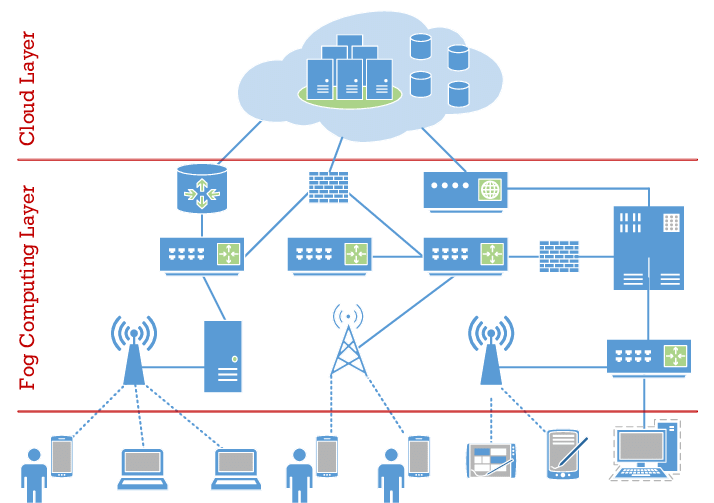
Frameworks that depend on fog computing for the most part produce high amounts of information or data, so more and more fog computing services and apps are arising. In addition, machine learning (ML), which is an essential area, has made significant advances in various areas of research including neuromorphic computing, decision-making computer graphics, natural language processing (NLP), speech recognition and robotics,. There are many studies that have been proposed to explore how ML can be used to solve fog calculation problems. As of late, an expanding pattern has been seen in reception of ML to give fog services and improve fog computing applications. for example, effective asset the board, security, alleviating inactivity,and traffic demonstrating and energy utilization. The current flow research shed light on introducing a diagram of the ML capacities in fog computing figuring territory or especially, we can say this paper proposes an AI way to deal with anticipating the remaining task at hand on fog nodes. The ML application for fog computing turns into a solid end-client and high layers administrations to increase significant investigation and more shrewd reactions for required assignments. We present an introduction to Fog computing, role and need of machine learning in fog computing, workload at the fog/edge nodes and finally machine learning approach to predict the workload.

# Introduction to Fog Computing:

## What is Fog Computing?

Fog computing and cloud computing are closely related to each other. So let's first describe what cloud computing is. Cloud computing was invented by Joseph Carl while he was working on a platform to connect people and data.

Cloud computing includes the processes like ICT(Information Communication Technologies) tasks, storing the results and data online using the internet. By using cloud computing, a business can use online hosted third party software or hardware. Because of the use of Cloud computing, it becomes easy to access information and computer resources from anywhere as per the availability of the internet. With the extensive availability of shared/pooled computing resources, cloud computing offers advantages over traditional on-site hosted services in terms of cost, speed and efficiency.



*[1] Layered Architecture of Fog Computing*

Although cloud computing is currently working fine, it is profoundly subject to the accessible bandwidth, which relies upon the limit of the network service provider. With billions of clients processing, sending and getting information in the cloud, the system turns out to be overloaded. Due to such heavy load and network latency issues cloud computing cannot be used for the applications where rapid action is required like patient monitoring in hospitals and pipe leakage monitoring in factories. For such applications fog computing comes into picture. Fog provides a computing layer near to end devices. The devices are connected directly to the fog layer and this reduces latency to a large extent which results in taking rapid actions.

Although fog computing provides many cloud computing facilities near the end users layer but still it can’t replace cloud computing. As for many applications like storage of large data fog can't be used. So fog computing is an expansion of cloud computing but not cloud at a lower level.

Instead of having devices traversing the network backbone infrastructure, Fog Computing allows devices to easily connect directly to their destination and do their connections and tasks at will. As a result, fog computing improves quality of service, reduces latency, and provides a more satisfying user experience.

## Characteristics of Fog Computing:

### **Low Latency**:

Due to the availability of the fog nodes with effective and keen end gadgets, the examination and age of information by these gadgets is quicker. This prompts a lower latency of information.

### **Heterogeneity:**

Fog computing is a significant heterogeneous framework since it can gather information from various sources. As a virtualized stage that gives end-client storage and different services, for example, networking, it goes about as an extension between end gadgets and conventional cloud computing centers.

### **Mobility:**

Many applications of fog computing need mobile devices for communication .This makes them suitable for techniques of mobility like LISP (Locator/ID Separation Protocol). The main work of LISP is to decouple the identity and location.

### **Scalability and Agility of Fog Node clusters:**

Because at cluster level it is versatile in nature, it could uphold the majority of the functions like network variations, data -load changes, and elastic compute.

### **The Dominance of Wireless Access:**

Although fog computing is widespread in wired environments. However, the wireless sensors, which are spread over large areas associated with IoT devices, place different demands and requirements related to analytics. Fog computing for wireless IoT access networks is also suitable for this.

## Advantages of Fog Computing:

### **Latency Reduction**

### The primary benefit of fog and edge computing is reduced latency. It is not necessary to send data to the cloud for processing as a part of the computation could be performed near to the data source for time-sensitive services.

### **Response Time Improvement:**

By reducing network latency, response time will be improved for the real time applications and will have a better user experience.

1. **Boosted Compliance**

Instead of moving data to the cloud, It can reside locally. This could increase compliance for particular business sectors.

1. **Overall growth in Efficiency and Speed**

In case there are a number of local user and IOT devices in which data is shared and local processing is allowed between them instead of using cloud services will enhance the overall efficiency and speed of service.

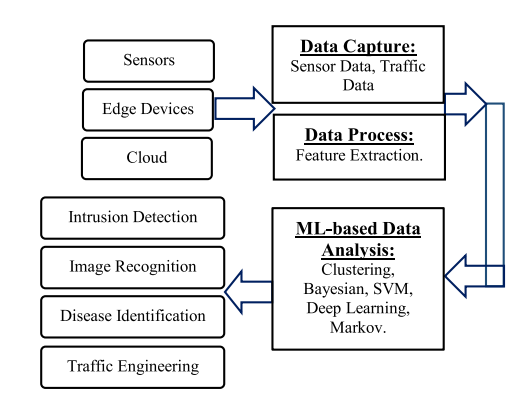
1. **Higher Data Privacy**

The processing of sensitive data is done locally and only a subset of data is being sent to the cloud for additional analytics.

# Fog Computing with Machine Learning:

In the modern age, use of computers and smart devices has increased drastically. Some examples are IoT devices and different sensors for temperature and light and many others. The amount of data generated by these devices is massive. At present, more and more organizations are shifting to cloud computing because of the advantages like accessibility and scalability. This shift requires dynamic information technology infrastructure. But due to the massive amount of data, it is not feasible to transfer all the data to the remote cloud that is generated. There might be several reasons for this. There can be storage, time and networking constraints. Here fog computing comes into the picture. In fog computing, we try to provide the cloud functionalities at some smaller level. This can help in reducing the response time of the application.

The data generated by IoT devices is massive. Sometimes there can be some incorrect or incomplete data. This data should also be removed to improve the results. Here we use Machine Learning to solve this problem. We can use ML models to remove these anomalies. The size of data that needs to be sent to the cloud can be reduced using some ML models.



*[1] Role of Machine learning in Cloud Computing*

## Uses of Machine Learning in Fog computing:

### **Computing Improvements**

Decentralised ML algorithms can be used to prevent sending the data to the cloud. We can use some data distribution algorithms to make the system resistant to the problem of networking connectivity. In this way, the data loss can be avoided or reduced during some network blackouts. ML supervised techniques are commonly used in healthcare and time-critical applications. On the other hand, unsupervised ML techniques are commonly used in different IoT applications, like traffic management and smart farming.

### **Decision Making**

By the use of directional mesh network framework, time-sensitive signalling data can be analyzed near the source with the help of different ML techniques. In the similar way, the fog devices can make some smart decisions regarding when the data should be sent to the cloud and when not to. This choice is accomplished with the help of low resource ML on fog devices that are used near smart telehealth devices. Generally Unsupervised ML models or techniques can be used to enhance the decision-making ability of fog computing where problems are related to computation of loads with limited resources and in some unexpected circumstances.

### **Prediction and Resource Provisioning**

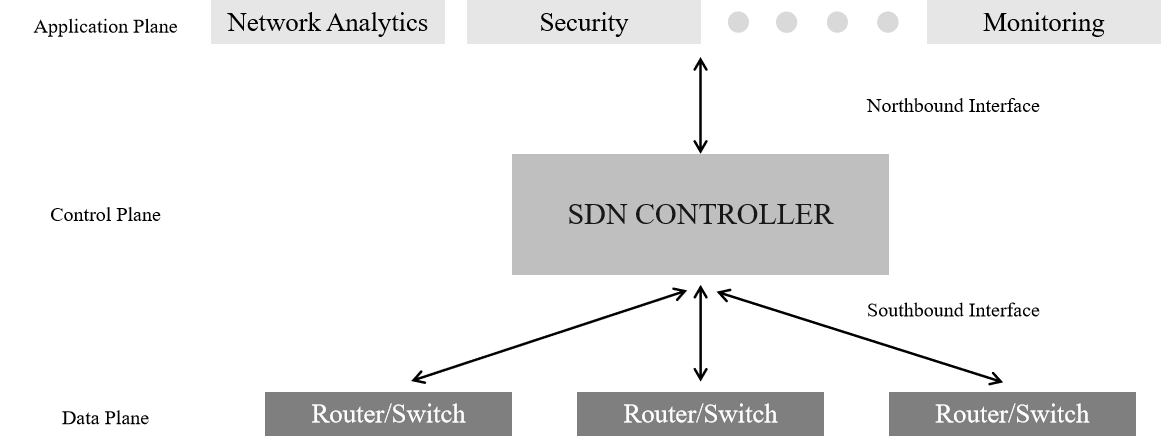
The resources required by applications can be analysed with the help of ML. By using resource provisioning we can improve processing speed and memory utilization. A model based on sophisticated ML algorithms can be used in multimedia fog computing for provisioning of resources. The primary functionality of the algorithm is forecasting the available resources in fog devices. This algorithm can also do some other tasks, such as checking the accuracy of the generated results and optimizing the limit of the similar job assignments. Some ML algorithms can be used in prediction of end to end delay like the time amount used from the processing to sending the data to the cloud, along with link utilization for image processing of different workloads.

## Fog Computing Security:

Fog computing is expected to be provided with reliable and secure services to the users using IoT devices and networks. Fog is a good location to identify which data requires what type of operations by analysing the data. By making the data anonymous, the system security can be improved. Cyber threats in distributed systems are high and increasing day by day. The developers generally prioritize supporting functional systems over incorporating security features first. This tendency also makes the possible cyber threats high. We can improve the query evaluation system’s efficiency by using an ensemble bag, neural network, boost, linear regression, and SVM on real and simulated datasets. These prediction models can facilitate the preservation of data privacy and elimination of the mean absolute error.

# SDN Enabled Networks for Fog Computing

SDN (Software-Defined Networking) is an emerging architecture that is cost-effective, dynamic, adaptable, and manageable, making it ideal for the applications that require high-bandwidth and dynamic in nature. This architecture separates the forwarding functions and network control, enabling the network control to be directly programmable and the underlying infrastructure to be abstracted for network services and applications.



SDN Architecture

## Network Quality Parameters:

1. Network Delay Components Model:

When we discuss the terms like quality, throughput and latency in networking, we do this in the context of the packet switching network. In this network, end-to-end delay means the total spent time for the L bit length data to reach from source node to the destination node.

for nodes from 1 to n.

1. Link Utilization Factor:

Another important parameter is link utilization in networking. It is used in proactively avoiding the violations of assessing Quality of Service (QoS). The link utilization parameter, b is as follows :

where dack is the total delay assuming the acknowledgment-frame of size 48-bit. is end- to-end delay and T is standard transmission delay calculating formula. Variable a is the ratio of

where is the rate of transmission and L is total no. of bits present in the file.

## Supervised Learning Models :

1. Linear Regression Model :

Linear regression is a **linear model**, for example a model that assumes a linear relationship between the single output variable (y) and input variables (x). More explicitly , A linear combination of the input variables (x) can determine y .

*y = B0 + B1.x*

1. Kernel-Ridge Regression(KRR) :

Linear regression was the first model to assess for regression tasks. However, a few highlights that are gathered in a fog computing environment might be profoundly associated or have collinearity, which can bring about the model being more defenseless to the over- fitting. The approach to cure the issue is to utilize the KRR(Kernel Ridge Regression). KRR is a non-parametric form of ridge regression.The point is to get familiar with a function in the space prompted by an individual kernel by limiting a squared loss with a squared standard regularization term.

The solution can be written as:

where K is the kernel matrix and α is the vector of weights in the space induced by the kernel.

1. Support Vector Regression (SVR):

Support Vector Regression is like KRR, in which it utilizes bit stunts. But still it has some principal dissimilarities. They utilize different loss functions. KRR uses squared error loss function. On the other hand Support Vector Regression utilizes ε-insensitive loss. Support Vector Regression is part of the Support Vector Machine (SVM) family.Support Vector Regression obtains similar characteristics related to learning. Observational outcomes have demonstrated that SVR will in general take more time to prepare than KRR however is quicker at making forecasts because of learning a sparse model.

One of the primary focal points of SVR is that its computational intricacy doesn't rely upon the dimensionality of the information space. Moreover, it has phenomenal speculation ability, with high forecast accuracy.

1. Gaussian Process Regression:

Gaussian process regression is a non-parametric, powerful Bayesian approach for regression issues that can be used in scenarios like exploitation and exploration.

GPR has few advantages, has the ability to provide uncertainty measurements on the predictions and works good on small datasets.

The purpose of using the Gaussian Process is crucial supposition that mapping of independent to dependent variables can’t be adequately caught by a single GP. GP utilizes lazy learning, that is used for delaying generalization of training data until the query has been made. The advantage of using the lazy learning method, i.e. instance - based learning, is that for each query, local approximation of target function is done.

These properties make lazy-learning generally helpful for huge datasets with several attributes which reflects that in an offloading environment, there is limited availability of features.

1. Random Forest Regression:

Random Forest is a group procedure capable of performing both classification and regression tasks with the use of numerous decision trees and a technique called Bootstrap and Aggregation,generally called bagging. The basic idea behind this is to determine the final output by combining multiple decision trees instead of depending on individual decision trees.

The base learning models of Random Forest are multiple decision trees.For every model,feature sampling and row sampling from dataset forming sample dataset for every model is performed randomly.This part is known as Bootstrap.

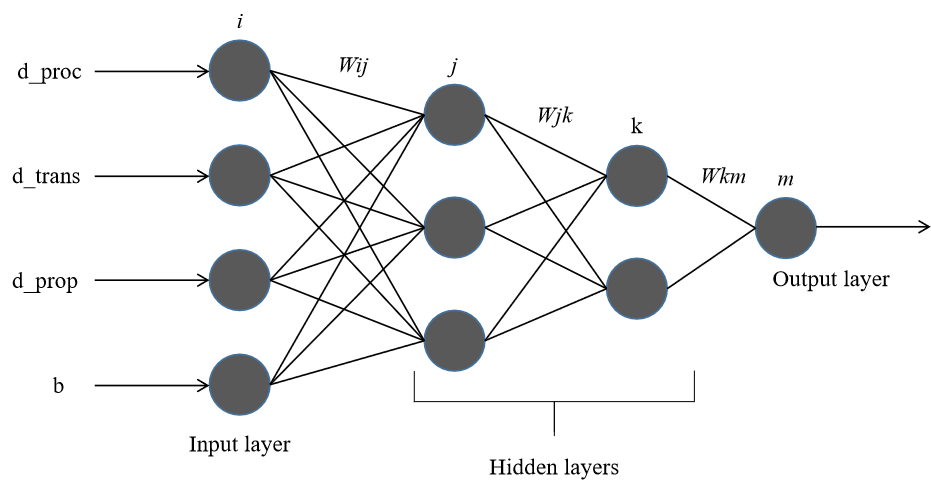
1. Multi-Layer Perceptron:

A **multilayer perceptron** (MLP) is a class of feedforward artificial neural networks (ANN). The term MLP is used ambiguously, sometimes loosely to *any* feedforward ANN, sometimes strictly to refer to networks composed of multiple layers of [perceptrons](https://en.wikipedia.org/wiki/Perceptron)

It has three hidden layers. For each hidden layer, there is a rectified linear unit activation function. Rectified linear unit(ReLU) is the most famous activation function that is used in deep learning because it is faster training, simpler and reduces the problem of vanishing gradients.

# Proposed Solution Approach

Workload at the fog or edge nodes can be predicted using the Artificial Neural Network(ANN). The ANN was trained by using Multi-Layer Perceptron(MLP) which was explained earlier in Supervised Learning Models. ANNs are the computational systems that simulate the neurons of the biological nervous system. In ANN, each layer is fully connected to the preceding layer by interconnection strengths and weights which can be seen in the diagram below[4]. The parameters used in the input layer are processing delay(dproc), transmission delay(dtrans), propagation delay(dprop) and link utilization parameter(b).



The solution or algorithm can be explained in following procedures:

1. Set the initial values of weights *Wij* and *Wjk*.
2. Calculate the outputs for all layers starting with the input layer.

*netj* = *WijXi*,  *j=1,2,3….J-1 , i=1,2,3….I*

,

*netk* = *WjkYj*,  *j=1,2,3….J-1 , k=1,2,3….K*

,

*f(x)* is sigmoid transfer function formula given by

*f(x) =*

1. Calculate the root mean square error(RMSE), Mean absolute error (MAE) to determine the performance of ANN.

RMSE =

MAE =

RMSE signifies error that gives difference between observed and modeled values.

1. We have to minimise the errors to get better results. So, there are some iterations required to get the minimum or sufficient error. When we reach the minimum error, iteration can be stopped.
2. Calculate learning error for every neuron in all layers:

= (- ) , *k=1,2,…,K,*

= ,  *j=1,2,..J-1 , k=1,2,...K*

K: total number of patterns

: desired output

: actual output

1. Update the weights Wij and Wjk

,

: learning rate

: momentum

: learning cycle

1. The all above procedures will be repeated until we find the desired value of error.

# References:

1. A Review of Fog Computing and Machine Learning: Concepts, Applications, Challenges, and Open issues

[*https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8869895*](https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8869895)

1. [*https://www.researchgate.net/figure/Three-layers-planes-of-the-SDN-architecture-This-presents-the-logic-of-an-entire-SDN\_fig1\_334683583*](https://www.researchgate.net/figure/Three-layers-planes-of-the-SDN-architecture-This-presents-the-logic-of-an-entire-SDN_fig1_334683583)
2. Predictive Analytics for Fog Computing using Machine Learning and GENI

[*https://www.researchgate.net/publication/325324436*](https://www.researchgate.net/publication/325324436)

1. The Use of Artificial Neural Network for Prediction of Dissolution Kinetics

*https://www.hindawi.com/journals/tswj/2014/194874/*

1. [*https://opennetworking.org/sdn-definition/*](https://opennetworking.org/sdn-definition/)