Vehicular Pollution Monitoring and Controlling using Fog Computing and Clustering Algorithm

Rajesh R.K.
Assistant Professor, Department Of Computer Applications
Mohandas College Of Engineering And Technology

Shijimol V.R.
Assistant Professor , Department Of Computer Applications
Mohandas College Of Engineering And Technology

Abstract—This paper proposes a solution for different pollution problems created by vehicular network. The vehicles emit poisonous gases which affects the entire human environment. Vehicular pollution has grown at an alarming rate due to growing urbanization in India. The air pollution from vehicles in urban areas, particularly in big cities, has become a serious problem. The pollution from vehicles has begun to tell through symptoms like cough, headache, nausea, irritation of eyes, various bronchial and visibility problems. The gas sensors implanted at the vehicles sense the pollution and actions are taken to inform the concerned people. This paper helps to monitor and control the poisonous gas emitted by vehicles.

Keywords - IoT, Client Server Computing, Cloud Computing, Fog Computing, Clustering mechanism, K-mean algorithm, Pillars of IoT, RFid Mechanisms.

I. INTRODUCTION

Human beings are affected with several diseases like heart disease, lung cancer, pneumonia, bronchitis due to air pollution**.** As per the report from World health Organization, air pollution is an important risk factor for multiple health conditions like skin infection, eye infection etc.So it is very important to monitor and control the air pollution. The best way to control air pollution is to monitor the exceeding levels of air pollutants and taking actions to control it. We have used several techniques to monitor the air pollution data. Vehicular pollution leads to instability, harmful and undesirable effects in the atmosphere. As the number of vehicles operated gets increased the air pollution has become a large area of concern. The primary pollutant is Carbon Monoxide which is a very poisonous gas. All these problems are identified through sensors implanted on the vehicles and data from these gas sensors is received and processed by fog node.

When the pollutant value is above the permissible level, strict and faster action should be taken to inform the pollution control board to take action. Here the vehicles are connected to fog node through Internet of Things.

II.INTERNET OF THINGS

The Internet is essentially a network of networks. Each of us connects to the Internet using a physical cable or through wireless media. Still more than 99% of things in the physical world are not connected to Internet. But with the Introduction of Internet of Every Thing, the scenario is going to change dramatically. Every physical device in the real world will be connected to Internet through RFID sensors, actuators etc. and these devices become intelligent also.

III. PILLARS OF INTERNET OF THINGS

There are four pillars for Internet of thing.
Things, Data, People, Process

Things: These are physical devices connected to Internet for intelligent decision making. The IoT will include all types of objects, including objects and devices that are not traditionally connected. In fact, Cisco estimates that 99 percent of physical objects will one day be connected. These objects contain embedded technology to interact with internal servers and the external environment. These objects are network-capable, and can communicate across a secure, reliable and available network platform.

Data: We collect data from things. It may be of different types based on the type of things. Data is a value assigned to anything that is around us. Data is everywhere. However, by itself, data can be rather meaningless. As we interpret the data, for example, by correlating or comparing, it becomes more useful. This useful data is now information. As this information is applied or understood it then becomes knowledge. In digital communication, data is represented as 1s and 0s.

People: They are the takers of Data. A large amount of data which no one can access, serves no one. Organizing that data and transforming it into useable information enables people to make better-informed decisions and take appropriate actions.

Process: This is an important pillar of Internet of Everything. Delivering the right information to right people at right time is very important. Processes are facilitating interactions between people, things, and data. Today, the IoE brings them all together by combining machine-to-machine (M2M), machine-to-people (M2P), and people-to-people (P2P) connections.

IV. SENSORS IN INTERNET OF EVERYTHING

A popular type of sensor uses radio frequency identification (RFID). RFID uses radio frequency electromagnetic fields to communicate information between small coded tags (RFID tags) and an RFID reader. Usually, RFID tags are used to identify and track what they are embedded into, such as a pet. Because the tags are small, they can be attached to virtually anything including clothing and cash. Some RFID tags carry no batteries. The energy required by the tag to transmit information is obtained from the electromagnetic signals that are sent by the RFID tag reader. The tag receives this signal and uses part of its energy to power the response.

V. COMPUTING MODELS

Client Server model: The primary method that businesses use to process data is a client-server model. This server collects IoT devices data from the client end usually a Router. Server does the required computation of the data. End users within an organization can store any number of files and documents on the file server, allowing end devices to conserve memory and processing power for use on local applications. By storing files on a central file server, other users within the organization can easily access these files, which allows for greater collaboration and sharing of information. Finally, with centralized services (such as file servers), organizations can also implement centralized security and backup procedures to protect those resources.

But in Today’s context, the client-server model is not always the most effective option. Because more devices are connected to Internet (IoT) from greater distances, having a centralized server may not be optimal. The devices which are farther away from the server may experience greater delays and more difficulties accessing the information. These changes in requirements for organizations and individuals have led to Cloud computing.

Cloud computing: In Cloud Computing model the Servers and Services are dispersed all over the globe in distributed data centers. Cloud computing allows IoT devices to access applications from servers located in the Cloud. In Cloud computing, data is synchronized across multiple servers, so that servers in one data center maintain the same information as servers in another location. Cloud computing has solved many problems of the traditional client-server model. Still Cloud computing is not the best option for delay-sensitive applications that require an immediate, local response. This led the IoT to use Fog Computing Model.

Fog Computing Model: It is also known as fogging, is a distributed *computing* infrastructure in which some application services are handled at the network edge in a smart device and some application services are handled in a remote data center in the *cloud*.

The significant challenge of Cloud Computing is connecting these devices and data centers where data are analyzed and processed. These devices can produce huge amounts of data. For example, in just 30 minutes a jet engine may produce 10 terabytes of data about its performance and condition. It would be inefficient to deliver all the data from IoT devices into the Cloud to be analyzed and then forward decisions back to the edge. Instead, some of the analysis work should take place at the edge, for example, on industrial-strength Cisco routers built to work in the field.

Fog computing creates a distributed computing infrastructure closer to the network edge that carries out easier tasks that require a quick response. It reduces the data burden on networks. It enhances resilience by allowing IoT devices to operate when network connections are lost. It also enhances security by keeping sensitive data from being transported beyond the edge where it is needed.

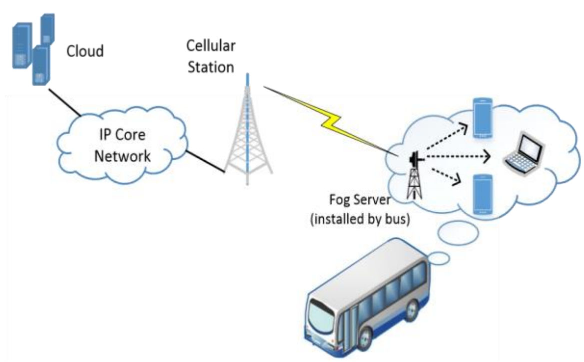

Figure 1: Fog computing Model

VI. CLUSTERING CONCEPT

The process of grouping a set of physical or abstract objects into classes of similar objects is called clustering. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters. It is a useful technique for the discovery of data distribution and patterns the underlying data

K-mean is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed apriori. The main idea is to define k centers, one for each cluster. These centers should be placed in a cunning way because different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest center. When no point is pending, the first step is completed and an early group age is done. At this point we need to re-calculate k new centroids as center of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new center. A loop has been generated. As a result of this loop we may notice that the k centers change their location step by step until no more changes are done or in other words centers do not move any more. Finally, this algorithm aims at minimizing an objective function know as squared error function given by:

$$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} (\|x_i - v_j\|)^2$$

Where,

'$||x_i - v_j||$' is the Euclidean distance between x_i and v_j.

'c_i' is the number of data points in i^{th} cluster.

'c' is the number of cluster centers.

The initial partitioning can be done in a variety of ways.

Dynamically Chosen: This method is good when the amount of data is expected to grow. The initial cluster means can simply be the first few items of data from the set. For instance, if the data will be grouped into 3 clusters, then the initial cluster means will be the first 3 items of data.

Randomly Chosen: Almost self-explanatory, the initial cluster means are randomly chosen values within the same range as the highest and lowest of the data values.

Choosing from Upper and Lower Bounds: Depending on the types of data in the set, the highest and lowest (or at least the extremities) of the data range are chosen as the initial cluster means. The example below uses this method.

VII. PROBLEM IDENTIFIED IN THE EXISTING SYSTEM

The focus of this paper is to monitor and control the air pollution. The best way to control the air pollution is monitoring whether the level of air pollutants is exceeding above a level or not. If it is exceeding the level action should be taken. There are so many reasons for this air pollution like gases emitted from factories, Vehicular exhaust where gases emitted by Transport vehicles and burning of plastic materials etc. here we are trying to monitor the Vehicular exhaustion. We need to implement Internet of Things with RFID sensors implanted on the Vehicles. The data are processed using fog nodes with gas sensors installed.

Initially data processing in Internet of Things Nodes at junctions were done by Cloud Computing model.

As we know cloud computing is an on demand service that shares a pool of resources over the network. In cloud computing data will be transmitted to the data center inside the core network and the result will be sent back to the end user after a series of processing. Suppose if we want to perform a real time processing of toxic pollution data the cloud may produce so many limitations. The latency and interoperability are the major issues to be solved in cloud computing. One of the most important issues in cloud computing is the performance overhead. Since fast accessing of data and resources is highly demanded especially provide timely feedback for emergency cases such as toxic pollution alert. In cloud environment the data i.e. the gas level read by the sensors nodes are transmitted to the data center inside the core network and the result will be sent back to the end user after a series of processing. By that time the pollution may have spread across the entire city making it vulnerable.

VIII. PROPOSED SOLUTION

Performance is the major concern in the field of cloud computing. Instead of storing the data in cloud fog computing can be used. Each vehicle will have a gas sensor which sense the pollution and this data is received by the fog node installed at traffic junction. This data is processed by the fog node and cluster the data using algorithms. The fog can be considered as an intermediate layer between the cloud and clients. Fog computing can provide elastic resources to large scale data process system without suffering from the drawback of cloud which is high latency. So here we aggregate data mined at fog nodes providing timely feedback for emergency case like toxic pollution alert. After a particular period of time the database will be taken and by using the K-mean clustering algorithm the data will be clustered. Here we form clusters based on the level of air pollutants emitted by vehicles. If the level of air pollutants emitted is above the permissible level or below the permissible level. Based on the level of air pollution processing should be done and appropriate action should be taken.

Air Quality Index Levels of Health Concern	Numerical Value	Meaning
Good	0-50	Air quality is considered satisfactory, and air pollution poses little or no risk.
Moderate	51-100	Air quality is acceptable; however, for some pollutants there may be a moderate health concern for a very small number of people who are unusually sensitive to air pollution.
Unhealthy for Sensitive Groups	101-150	Members of sensitive groups may experience health effects. The general public is not likely to be affected.
Unhealthy	151-200	Everyone may begin to experience health effects; members of sensitive groups may experience more serious health effects.
Very Unhealthy	201-300	Health alert: everyone may experience more serious health effects.
Hazardous	> 300	Health warnings of emergency conditions. The entire population is more likely to be affected.

Figure 2: The details about level of gas pollution and its consequences.

Algorithmic steps for k-means clustering

Let $X = \{x_1,x_2,x_3,\ldots\ldots,x_n\}$ be the set of data points and $V = \{v_1,v_2,\ldots\ldots,v_c\}$ be the set of centers.

1) Randomly select *'c'* cluster centers.

2) Calculate the distance between each data point and cluster centers.

3) Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers.

4) Recalculate the new cluster center using:

$$v_i = (1/c_i) \sum_{j=1}^{c_i} x_i$$

where, 'c_i' represents the number of data points in i^{th} cluster.

5) Recalculate the distance between each data point and new obtained cluster centers.

6) If no data point was reassigned then stop, otherwise repeat from step 3).

Advantages

1) Fast, robust and easier to understand.

2) Relatively efficient: O(tknd), where n is # objects, k is # clusters, d is # dimension of each object, and t is # iterations. Normally, k, t, d < n.

3) Gives best result when data set are distinct or well separated from each other.

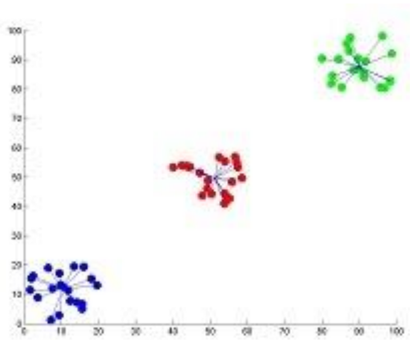

Figure 3: Showing the result of k-means for *'N'* = 60 and *'c'* =3

As a simple illustration of a k-means algorithm, consider the following data set consisting of the ppm levels of a particular city and we need to make 2 clusters.

Ppm level
10
400
100
300
360
25
560
250

Figure 4: ppm level indicating pollutants in a junction

This data set is to be grouped into two clusters. As a first step randomly partitioning the dataset like this.

	Ppm level	Mean Vector (centroid)
cluster 1	10	133.75
	25	
	100	
	400	
cluster 2	300	367.5
	360	
	250	
	560	

Figure 5: Clusters of data after the first iteration

The remaining cluster values are now examined in sequence and allocated to the cluster to which they are closest, in terms of Euclidean distance to the cluster mean. The mean vector is recalculated each time a new member is added. This leads to the following series of steps:

Now the initial partition has changed, and the two clusters at this stage having the following characteristics:

	Ppm level	Mean Vector (centroid)
cluster 1	10	45
	25	
	100	
cluster 2	300	374
	360	
	400	
	560	
	250	

Figure 6: Clusters of data after the second iteration

If we repeat the same process then there is no more changes .So we are getting 2 clusters like this. The first one contain the permissible ppm levels and the second cluster having the ppm levels above the limit. We pass on this information to Pollution control board or police so that they may catch the vehicle which caused the pollution.

IX.CONCLUSION

Thus the vehicular gas or pollutants are identified and clusters are formed. The cluster which has gas level below 150 ppm means that no action is taken and if the ppm level is above 150, strict action should be taken by informing the Pollution control Board. As the processing and analysis done at the fog layer we can improve performance and efficiency. Fog computing creates a distributed computing infrastructure closer to the network edge that carries out easier tasks that require a quick response. It reduces the data burden on networks. By using K-means algorithm the analysis can be done by taking the data as cluster. So that we can give a fast response for each query. Thus the environment is saved from thousands of pollution.

REFERENCES

[1] Fog computing – Focusing on Mobile users at the edge by Tom H. Luan, Longxiang Gao, Yang Xiang and Zhi Li, Limin Sun on 01815v1 [cs.NI] 6 Feb 2015

[2] Fog Computing: A New Concept To Minimize The Attacks And To Provide Security In Cloud Computing Environment By Sonali Khairnar1, Dhanashree Borkar2 1M.E, Department of CSE, JSPM's Imperial college of Engineering and Research, Maharashtra, India 2B.E, Department of CSE, JSPM's Imperial college of Engineering and Research, Maharashtra, India

[3] Computing at the Mobile Edge: Designing Elastic Android Applications for Computation Offloading Gabriel Orsini, Dirk Bade, Winfried Lamersdorf Distributed Systems Group, Department of Computer Science University of Hamburg, Germany

[4] Help Your Mobile Applications with Fog Computing Mohammed A. Hassan, Mengbai Xiao, Qi Wei and Songqing Chen mohammeh@netapp.com, NetApp Inc.

[5] Internet of Things : Cisco certification course link: https://1299635.netacad.com/courses/338734

[6] An Efficient k-means Clustering Algorithm: Analysis and Implementation by Tapas Kanungo, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman and Angela Y. Wu.

[7] k-means clustering by ke chen.

[8] Improve Performance of clustering on Cloud Datasets using improved Agglomerative CURE Hierarchical Algorithm Mrs. Parekh Madhuri H, Prof.Ishan K.Rajani

[9] Intelligent Pollution Monitoring Using Wireless Sensor Networks By Gokulram.K1 , Dhakshinamoorthy.T2 1 P.G Scholar, Electrical and Electronics Engineering, Kumaraguru College of Technology, Tamilnadhu, India 2 Asst.Professor, Electrical and Electronics Engineering, Kumaraguru College of Technology, Tamilnadhu, India