International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 12, Number 10 (2019), pp. 1739-1750

Machines' Fault Detection and Tolerance Using Big Data Management

Salaheddin J. Juneidi †

Palestine Technical University Khadoorie/ Aroub,
Hebron, West Bank Palestine.

ORCID: 0000-0002-1983-5484

Abstract

Internet of Things (IoT) is one of significant topics in business and in research field. Modern societies are adopting IoT as a necessity rather than a luxury in their daily life style. IoT defined as devices and machines are connected to data repositories which mainly reside in the Cloud. One of the most challenging issues related to IoT environments is the huge amount of data, referred to as Big Data. Data generated from IoT environments are generally stored and processed in cloud infrastructure, then it consumed (in analyzed forms) as input to feed the same IoT environments. IoT and Big Data are two faces of the same coin. This article discusses IoT aspects related to big data in cloud, proposes a Big Data Management Center (BDMC) that utilizes accumulated big data transmitted form machines sensors .BDMC can anticipate faults that might occur in machines, besides it can decide which fault can be tolerated and which fault resulted in severe damage in that machine. This prediction will give us advantage in advance to overcome problems before they occur. This article will also go through important aspects relate to IoT such as "storage as cloud", "wireless connectivity" and data storage using Hadoop format. Finally, this article presents a case study on fault detection on real vehicles tests, then experimental results is provided.

Keywords: Sensors, BDMC, Fault detection, Machine Learning

1. INTRODUCTION AND BACKGROUND

Big data management and utilization is becoming a rising demand in business and modern life style,[1] researches have considered this topic as one of the most significant topics in research fields, big data as expression comes from the fact that magnificent amount of data accumulated and transmitted from running machines' sensors or smart programmable devices [2]. Modern technology categorized with abundant hi-tech resources that has very much influence on evolving this research field. Business demands are getting higher to adopt Internet of Things, IoT is basically making things like machines, devices, and even home appliances continuously data connected. The huge amount of data transmitted from those devices are used it to supervise and to control their performance. Companies start to use the collected data from devices to improve competence and customer satisfaction. Data is accumulated in data base centers, named as cloud, connected devices are typically given a name of smart device, each smart device is connected to cloud eventually through wireless network. The continuous transition of data from smart devices to cloud is accumulated in a huge amount of data that needs to be managed and analyzed for better utilization.

Normally machines have sensors in their different mechanical parts, those sensors are supervising and controlling the machines performance. For running machines, we need to have full view on their performance, also to anticipate for faults that might occurred in future. Then determine which fault can be tolerated and which is sever that might damage it's related mechanical part.

This research article will be helpful for researchers and practitioners who interested in utilizing big data produced by machine sensors, this article will explain main elements of machine fault detection and tolerance. The introduction section provides related issues concerning IoT like cloud computing, and then it discusses wireless networking and connections. Section 2 is presenting the current status and problem definition of vehicle repairing without big data. Section 3 is dealing with methods of big data collection and management in IoT. Section 4 is discussing big data storage and distributed file system. Section 5 is presenting big data general algorithms for prediction and detection. Section 6 is deliberating data visualization in a real case study and applying methods of fault detection and tolerance. Finally, section 7 concludes the article.

1.1 Cloud Computing

The *cloud* is a Big Data storage space for keeping and exchanging data with smart devices and machines through sensors and special data busses. As depicted in Figure 1, smart devices and appliances are connected to the Cloud computing system; data is transferred to the Cloud is processed by software systems. It is important for the Cloud infrastructure to have full access, monitoring, management and configuration capabilities on smart devices that are connected with. This is viable through a Communication Bridge Module (CBM).To be able to to process and to manage big data in the cloud, we need to specify the amount and the arrangement of transferred data to and from the smart devices. Depending on that, we choose one of the most suitable smart bus to convey data input/output to/from cloud. The modern trend in manufacturing of any device or modern machine is to engineer it as 'smart' as much as possible [2]. Customer service is the main objective of companies to gain competitive edge, generally companies perform this service by keepingtrack of history fordevice or machine serviceto gain customer satisfaction.

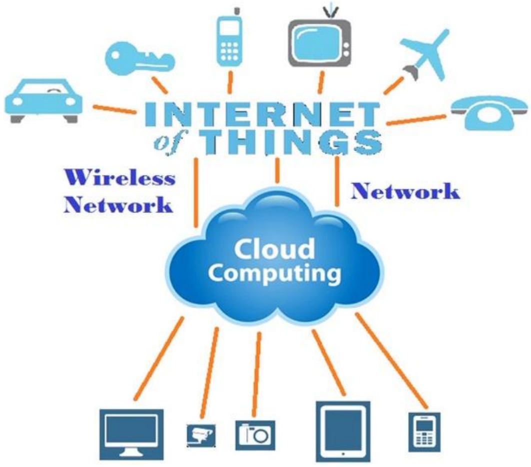


Figure 1. IoT and smart systems' main elements

1.2 Wireless networking

Wireless networking is one of the most vital factors in IoT, since, as part of their functionality requirements, devices must be continuously connected to the Cloud. Connections are set in various ranges, starting from meters (around buildings) to satellite connections. This paper will not go through technical issues of wireless connections but will discuss the trend worldwide.IoT, as showed Figure 2. The common for of connection in IoT is wireless point to point (P2P) protocols from devices and machines, connected to hybrid gateways and WAN gateways to collect data from different kinds of wireless network, then transmitted securely to the cloud storage, mainly the source is smart devices and machines through sensors and destination is the cloud system. Basically, the cloud defined as [3]: it is the case of servers with high connectivity, using specific wireless connection standards. The wireless technology is now advanced in fast paces, which provides the users the capability to fully use their networking standard by approving several characteristics. There are approximately 14 standards which are utilizing the connections for the IoT which well-known enabling technologies of are today. RFID and near-field communication, Li-Fi, Optical tags and quick response codes, Bluetooth low energy, Low energy wireless IP networks, ZigBee, Z-Wave, Thread, LTE-Advanced, WiFi-Direct, HaLow, HomePlug, Ethernet, MoCA are some of the commonly used enabling technologies [4], [5], and [6]. These wireless networks are various, they are vary according to specific characteristics such as network standards, connection protocols and the range they cover.

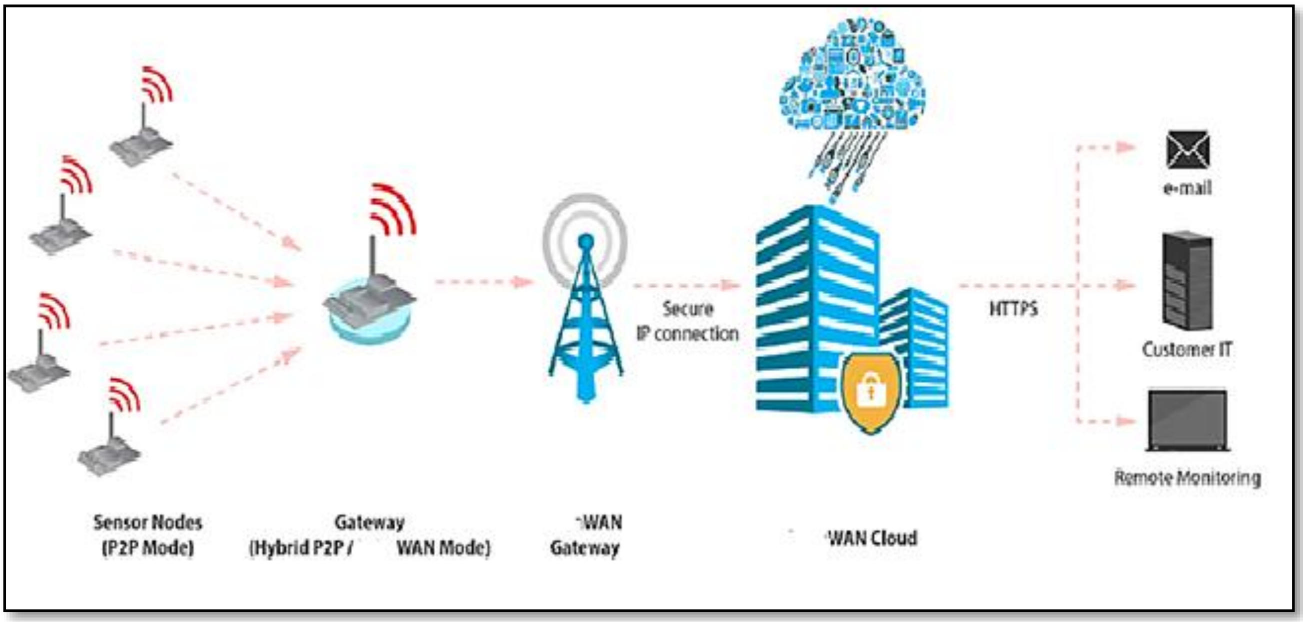


Figure 2. IoT and its Connection from smart machines to cloud

1.3 Big Data Management

One of the most important issues in IoT, perhaps, is the efficient management of big data. While smart devices are in operation, enormous amount of data is accumulated in the cloud. Big data theory is taking advantage of this data accumulated, which can be done via data statistical and analytical approaches, for the purpose of making data useful in a strategic manner. Big data management concerns with the utilization of all the pieces of data collected from continuously operating smart devices and machines within an IoT environment [7].

Table 1. Comparison between big data and traditional data

Components	Big data	Traditional Data
Queries	Largely Abandoned SQL	Traditional SQL
Architecture	Distributed	Centralized
Data Types	Structured,Semi-Structured and Unstructured	Structured
Data Model	No schema	Fixed Schema
Data Relationship	Unknown or complex Relationships	Known Relationship
Data volume	Petabytes or Exabytes	Terabytes
Data Traffic	More	Less
Data Integrity	Less	High

Big Data is differ from traditional data in various aspects , the container of traditional data is data base system .Normally big data is kept in its primitive and raw form, big data is lacking to uniform structure or static relations, big data is natively distributed , that kept in file fields that can be used instantly or for statistical reasons later on, SQL queries cannot be answered from big data , even data base principles like integrity and consistency are not main principle in big data , Table 1 gives conclusive comparisons between big data and tradition data that kept in relational data base. Big data main characterizes are [8]:

1. ***Volume*** *:* Volume refers to amount of data. Volume of data stored in enterprise repositories have grown from megabytes and gigabytes to petabytes. 40 zetabytes of data will be created by 2020 which is 300 times from 2005. The social networking sites existing are themselves producing data in order of terabytes every day and this amount of data is definitely difficult to be handled using the existing traditional system.

2. ***Velocity*** *:* Velocity in Big Data is a concept which deals with the speed of the data comingfrom various sources and the speed at which data moves around.

3. ***Variety:*** Big Data is not always data and it is not always easy to put big data into a relational database.

4. ***Veracity:*** When we are dealing with a high volume, velocity and variety of data, it is not possible that allof the data is going to be 100% correct, there will be dirty data. The quality of data being captured can vary greatly. The data accuracy of analysis depends on the veracity of the source data.

5. ***Value:*** Value is the most important aspect in the big data. Though, the potential value of the Big Data is huge. It is all well and good having access to big data but unless we can turn it into value it is become useless.

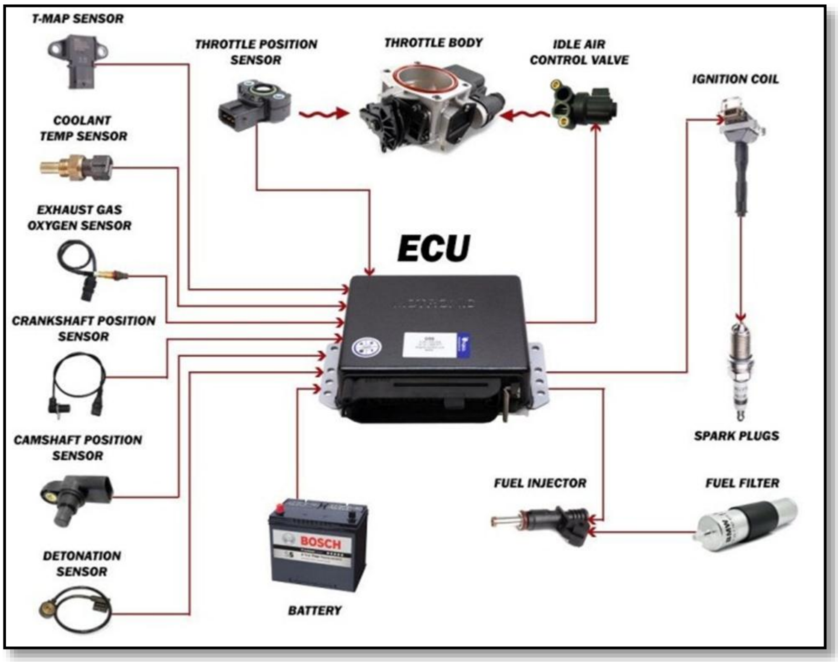


Figure 3: Vehicle's Engine Control Unit (ECU) with connected sensors

2. CURRENT STATUS AND PROBLEM DEFINITION

In automobile industry were compotator are very furious, and each of which are looking for a dominant competitive edge. One of the most competing factors for a manufacturer is customer satisfaction and service. Vehicles in general have many data sensor that are always reads data from different vehicle parts, such as engine behavior. Engine Control Unit ECU that has data inputs from various sensors installed on different vehicle parts, data concerns with checking performance of each part, ECU has also Outputs that varies according to sensors check input. The outputs and normally called *actuator* may produce some alarm voices or alarm lamp on drivers front panel, or might for worst condition make action to stop the machine before general failure happened . Figure 3 shows ECU and different sensors that connected with it, each sensor send data to the ECU for taking action accordingly. As shown in Figure 3, power is supplied to ECU directly from vehicle battery , the number of sensors in a modern vehicle may reaches to 200 according to the vehicle kinds and its supporting accessories, average vehicle sensors ranges between 60 to 100 [9] , Luxury vehicles have even much more sensors. The actuator is the acting part of ECU depending on input sensor readings, normally the actuator is responding to specific defect in the vehicle parts like oil range temperature of engine or even open door while driving. All vehicle defaults can be depicted by sensors, the mechanics can identify defaults according to AutoData application, this application is connected to actuator to read outputs and accordingly identify where is the damage , then the mechanics in garage can define the cause of error and the defected parts of engine that need to be fixed or replaced , a screenshot of this application is shown in Figure 4 , this application shows the current status of the vehicle engine and where are the defaults according to error codes given by sensors . according to mechanics some time the problem is very simple such as misconnection parts with ECU that affect some auto part, and this will be easy fixed , other cases may be more serious problems that need to repair and change some parts , in both cases the vehicle owner will call vehicle towing to handle vehicle problem in mechanic premises.

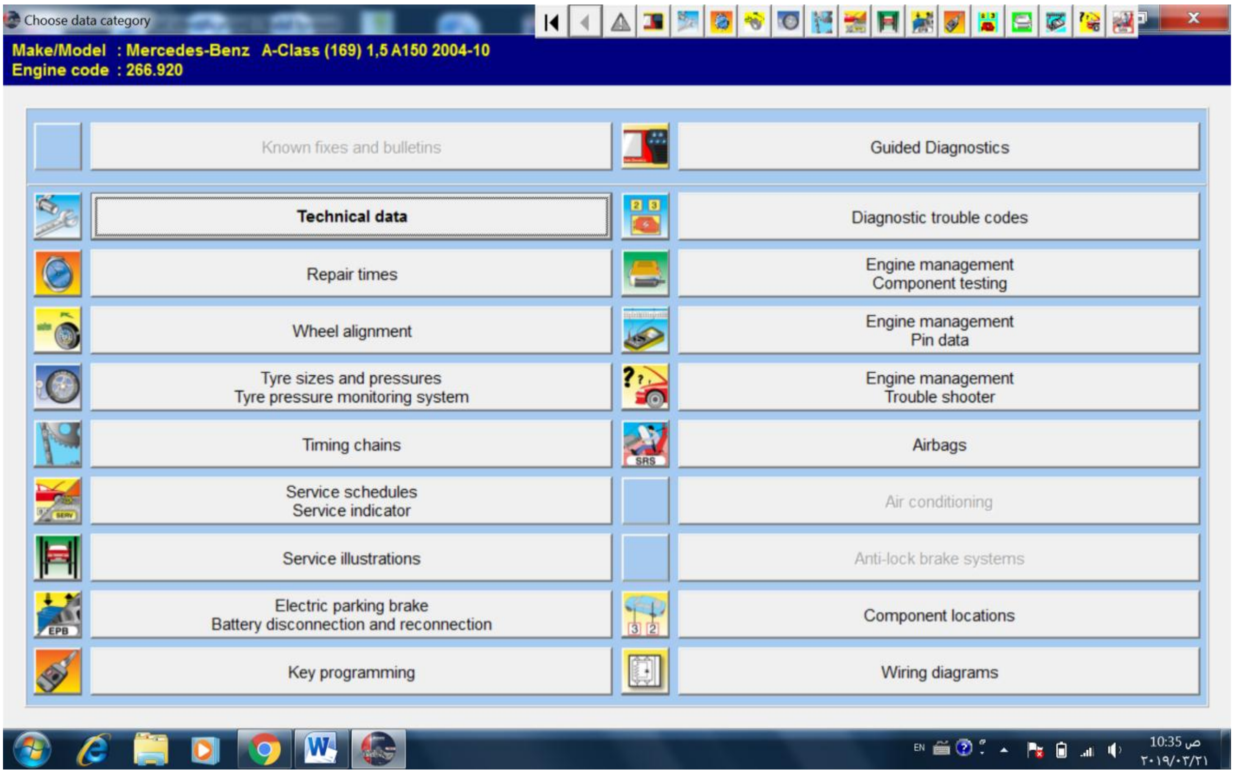


Figure 4: Software Auto Data to check errors and faults occurs in Motor Vehicles

This is the current situation in most cases of vehicle failure, this article proposes a better solution by using big data collected from ECU, instead of waiting until problem or failure occurs , we can anticipate the failure and do repairs before this failure get worsen and even damage other vehicle parts. This is the main theme of fault detection that presented in this article

3. BIG DATA MANAGEMENT FOR IoT:

In the previous section we have discussed the amount of sensors for each vehicle running on road, actually , periodically collection of data produced by these sensors needs remarkable amount of storage for just single vehicle, now imagine how much storage needed to collect and store data produces by amount of the cars in one geographical region , or even more how much storage we need to collect the sensors reading from vehicles in one country , that's why we need to use methods to collect , utilize and manage big data , cloud computing is very close to big data theory that's the data is stored in cloud , which provides a massive storage can be distributed in larger that district or geographical region. Normally, big data is collected from sensors on smart machines and devices, this data is transmitted through smart wireless busses to reach its destination storage of the Cloud see Figure 4 ., this article proposes multitasking system for collecting, visualizing and managing big data, Big Data Management Center (BDMC) is mainly responsible of receiving data from smart machines to the cloud , BDMC provides full surveillance of running machines, as enormous data is transmitted from smart machines to cloud supported by a real application provided later on in this article, this center deals with many variables concerns machines , these variables is determined according to the nature of the smart machine main functionality :.As soon as data get into the cloud it will be managed according to related readings from different sensors on machine, this management process committed by BDMC . The main management's tasks that take place on big data would be the following:

1. Data categorization and visualization: put each chunk of data received into categories, based on the authentication factor, example related sensors , logs, time, interrupts, faults, etc.
2. Sorting and searching tools: algorithms for storing data to make it easier for retrieving data with the least time of job accomplishment, especially for real time systems where time is very critical.
3. Predicting: to evaluate the lost or unclear data (noisy) and predict its actual value.

We notice in Figure 5, BDMC is distributed system as part of the cloud , connected with sensors with through wireless, collect data in its raw format , process data , and make decision about fault detection , then it will call or text the driver about serious faults that might detected though his / her smart phone.

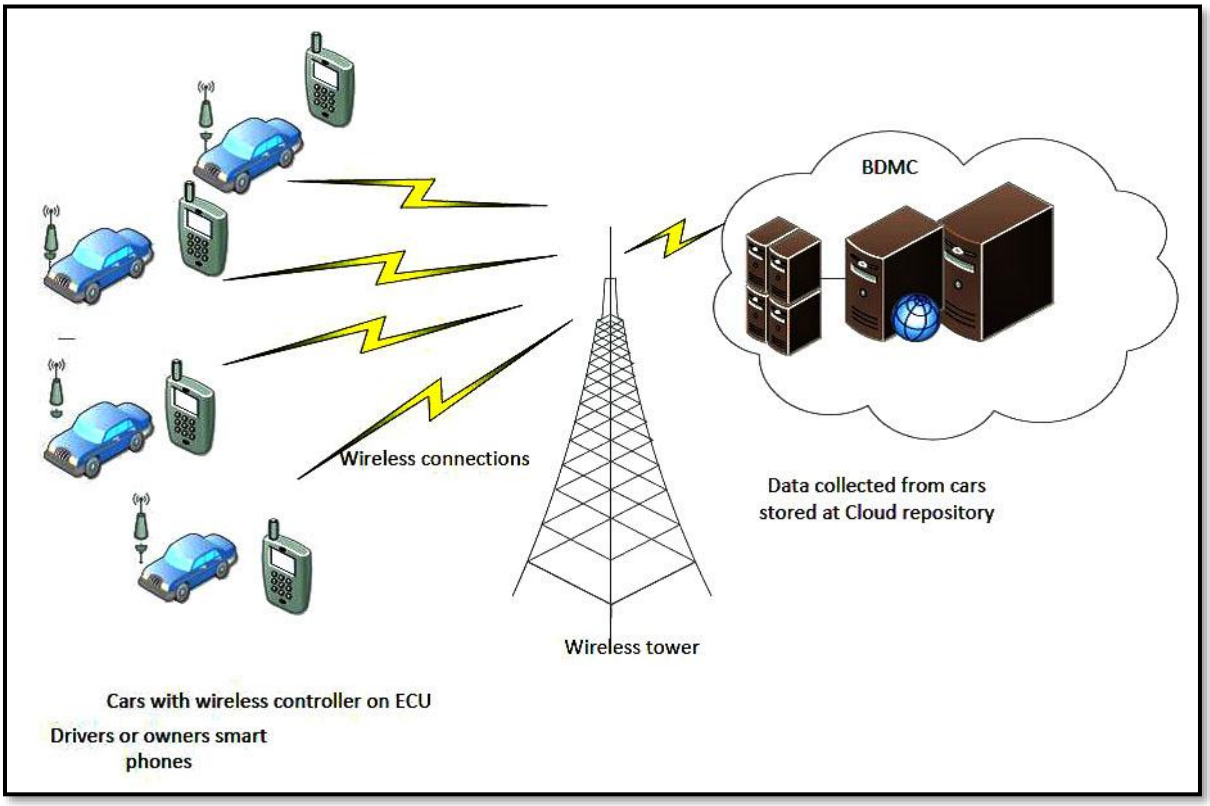


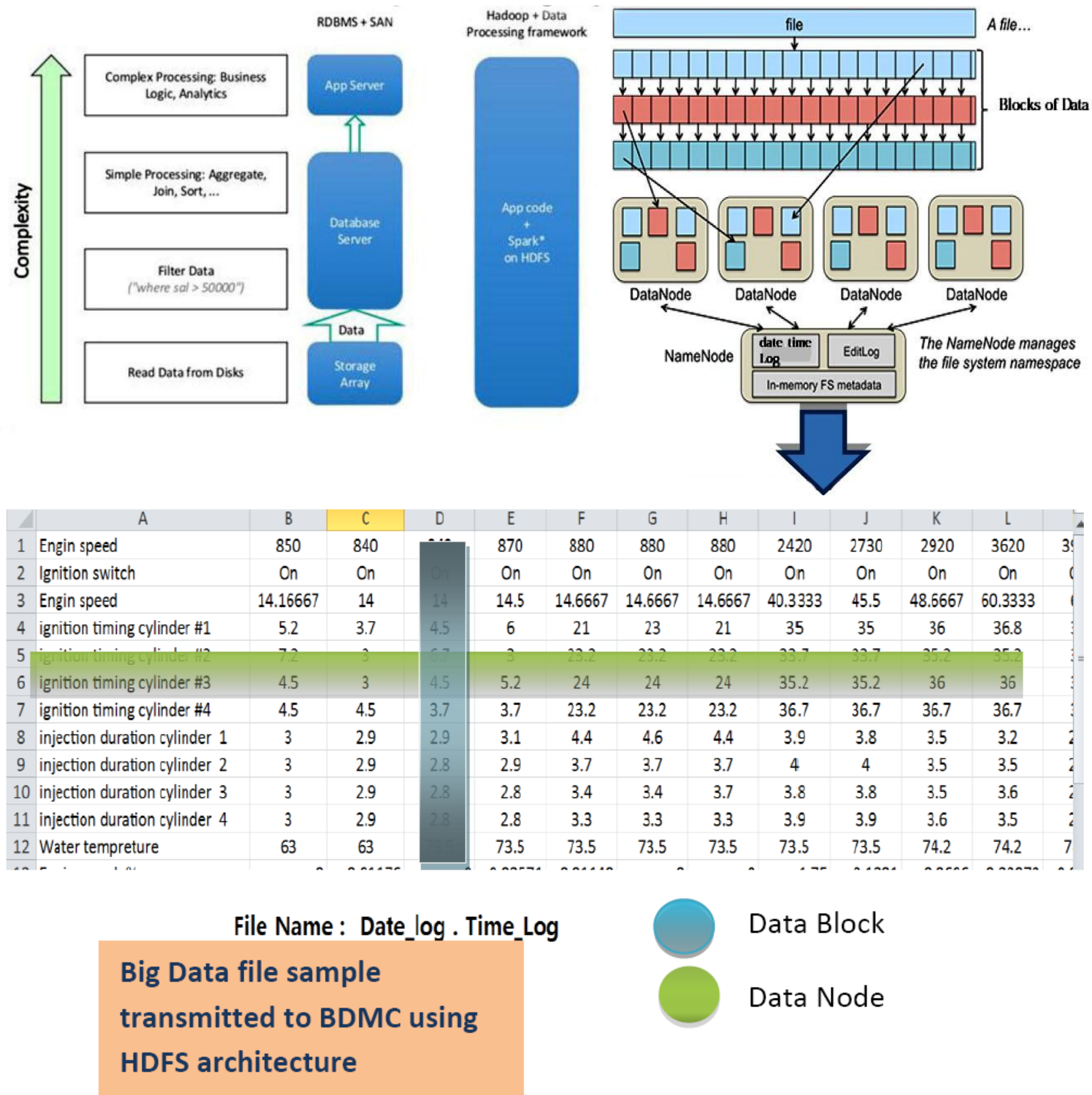
Figure 5: Vehicles wireless connection to Big Data Management Center BDMC

4. DATA MANAGEMENT AND VISUALIZATION

In this section we will define the software and architecture to support BDMC , from the characteristics of big data listed above in previous section , this article considers Hadoop [10] to represent this architecture, Hadoop (Highly Archived Distributed Object Oriented Programming) It is an Open source Java Framework technology helps to store , access big data files. It provides large resources for big data in a distributed fashion at less cost, in addition it has high degree of fault tolerance and high scalability [11]. Traditionally, relational databases .management systems are used to manage data. But, in this article we collect data of vehicle sensors in Hadoop architecture. We are not interested in allocating data using relational data base or into organized tables with the complication of normalization of database theory. From big data characteristics provided in table 1, Hadoop architecture is seems to be the most suitable container of big data from the following facts:

1- It deals with horizontally cluster scale of data using plain file system.
2- It can process simi-structured or even unstructured data
3- It provides scalable amount of storage of distributed computing by file system imaging.

Table 1 depicts the comparison between Relational Data Base Management System RDBM using Storage Area Network SAN, and Hadoop Distributed File System HDFS , we can see big data storage and processing are committed in the same level of execution as soon as the raw data file collected from machines sensors , unlike relational data base file system has non-structured or semi-structure data , Hadoop with spark system has no further arrangements in tables or transfer data to other hierarchy structures but data is stored and processed as soon as it arrives without any bottle neck transformation, and no transfer of data from one unit to another (storage-server- application) as what has happened in data base systems , data storage and manipulation in HDFS is simple and occurs at the same level of abstraction in Figure 6 we see that the data is collected in files with elementary structure called Blocks of data with no data structure and without related tables complications, then it distributed into Data Node , each node contains naturally related data , this relation may represent the time the data is read, or even user define equation that define this relationship .Final part of Figure 5 is very small instance sample of file that contains data received from vehicle sensors , the relation between these sensors is that they are important to represent the engine performance and default detection.



	A	B	C	D	E	F	G	H	I	J	K	L
1	Engin speed	850	840	[illegible]	870	880	880	880	2420	2730	2920	3620
2	Ignition switch	On	On	On	On	On	On	On	On	On	On	On
3	Engin speed	14.16667	14	14	14.5	14.6667	14.6667	14.6667	40.3333	45.5	48.6667	60.3333
4	ignition timing cylinder #1	5.2	3.7	4.5	6	21	23	21	35	35	36	36.8
5	ignition timing cylinder #2	7.2	3	6.7	3	23.2	23.2	23.2	33.7	33.7	35.2	35.2
6	ignition timing cylinder #3	4.5	3	4.5	5.2	24	24	24	35.2	35.2	36	36
7	ignition timing cylinder #4	4.5	4.5	3.7	3.7	23.2	23.2	23.2	36.7	36.7	36.7	36.7
8	injection duration cylinder 1	3	2.9	2.9	3.1	4.4	4.6	4.4	3.9	3.8	3.5	3.2
9	injection duration cylinder 2	3	2.9	2.8	2.9	3.7	3.7	3.7	4	4	3.5	3.5
10	injection duration cylinder 3	3	2.9	2.8	2.8	3.4	3.4	3.7	3.8	3.8	3.5	3.6
11	injection duration cylinder 4	3	2.9	2.8	2.8	3.3	3.3	3.3	3.9	3.9	3.6	3.5
12	Water tempreture	63	63	[illegible]	73.5	73.5	73.5	73.5	73.5	73.5	74.2	74.2

Figure 6: Comparison between data architecture in Relational Database Management System and Hadoop HDFS

5. DATA ANALYSIS METHODS

As we see from previous section data is transmitted from vehicle sensors , this data is received through wireless network into distributed files. These files basically have raw data accumulated each time of transmission , the period of transmission is set according to the system requirements , the file is saved with name to show the date and time logs , in HDFS each file has blocks and nodes , normally block is one column of data is a certain instant data of sensors received at a certainperiod of time, and the row is normally the data of the same sensor received in different period of times . at this stage , the big data file is ready for processing and analyzing , this article will focus on two issues regarding big data processing [12,13]:

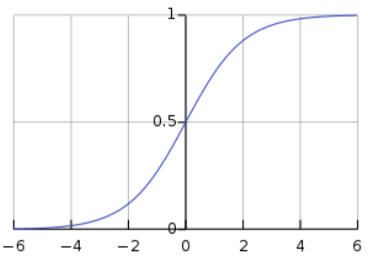
1. **Predictive analytics** : which uses data history , to produce activity , or define forecast behavior or trend, and may used to reach a decision ,normally this kind of data processing uses statistical analysis to define an algorithmic or model for application, BDMC , which can be done by:

a. *Logistic regression:* Logistic regression is a statistical analysis method used to predict a data value based on prior observations of a data set. A logistic regression model predicts a dependent data variable by analyzing the relationship between one or more existing independent variables, we going to use logical regression in default detection , after we define the curve for related sensors readings. the general algorithm of logical regression is presented as :

$$P = \frac{e^{a+bX}}{1+e^{a+bX}}$$

Or

$$P = \frac{1}{1+e^{-(a+bX)}}$$



Where P: is the value of prediction
X: is independent variable
a,b : are the parameters of the model

b. *Decision tree:* A decision tree is a tree where each node represents a feature (attribute) or sensor reading , each link(branch) represents a decision(rule) or relation between sensors and each leaf represents an outcome (categorical or continues value), decision tree can be used in BDMC to check if sensor readings go out of its rang , the result here is whether the readings yield to a fault or not. In general the decision tree is given by this example of whether forecast :

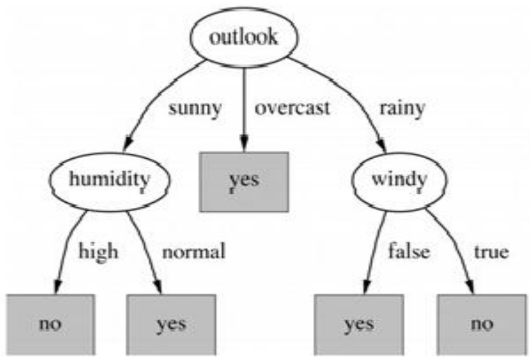
The objective is to determine yes or no for two variables. There can be 4 combinations.

Gini Index for Binary Target variable is

=1-P^2(Target=0) P^2(Target=1)

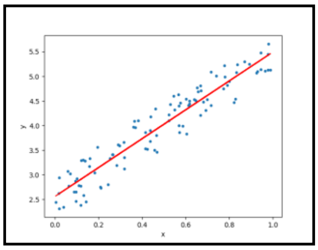
$$= 1 - \sum_{t=0}^{t=1} P_t^2$$

Gini Index is our cost function used to evaluate splits in the dataset.



2. **Machine Learning** :Transmitted data is heavily depending on wireless connection , sensors readings are transmitting in periodical bases to the BDMS , sometimes data connection face troubles though wireless network by noisy or unclear data transmission , as shown in Figure 7, if the data read from one sensor is not clear we will apply one of the machine learning algorithm to determine the most accurate reads for the missconnected sensor data, normally in BDMC the noisy algorithm value is considering between 2 and 4 consecutive time periods before the corrupted reading and after the reading becomes good and clear , the anticipate value algorithm is depending on the time consecutive periods for readings form sensors and the stability of the other related readings . as in Figure 7 we apply mainly two machine learning algorithm to anticipate bad readings:

a. *Linear Regression :* one of the simplest machine learning algorithm's, it used statistical view in a set of variables , determine the equation of two or more related variable, according to linear equation we can determine some of missing values , this algorithm can be used to set noisy and unclear sensors readings . We can determine missed points according the linear relations and equation with other points around the linear curve.



The linear regression model can be represented by the following equation:

$$Y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + ... + \theta_n x_n$$

Y is the predicted value

$X_1, X_2 ... X_n$: Dependent variables

θ_0 is the bias term

$\theta_1,\dots,\theta$ are the model parameters

$$Y = \theta^T x$$

where

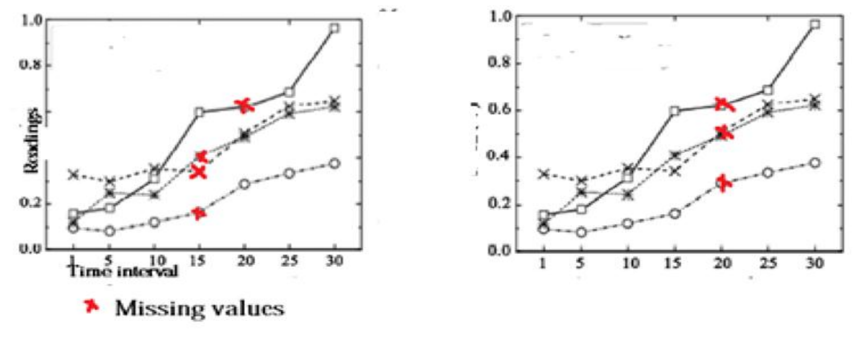
θ is the model's parameter vector including the bias term θ_0

x is the feature vector with x_0 =1

b. **K-Nearest Neighbor:** This machine learning algorithm is non-parametric algorithm used for classification and regression ,basically determine noise or misconnection ambiguous sensor readings from the closest sensor reading before and after if , as presented in Figure 9 , the corrupted or miss value in in red cross. BDMC takes the average of two prior and post reading from the same sensor .the general form algorithm of K-NN is :

$$A = \frac{1}{n}\sum_{i=1}^{n} a_i = \frac{a_1 + a_2 + \cdots + a_n}{n}$$

Where A is the average , a is the reading values, and n is the number of values



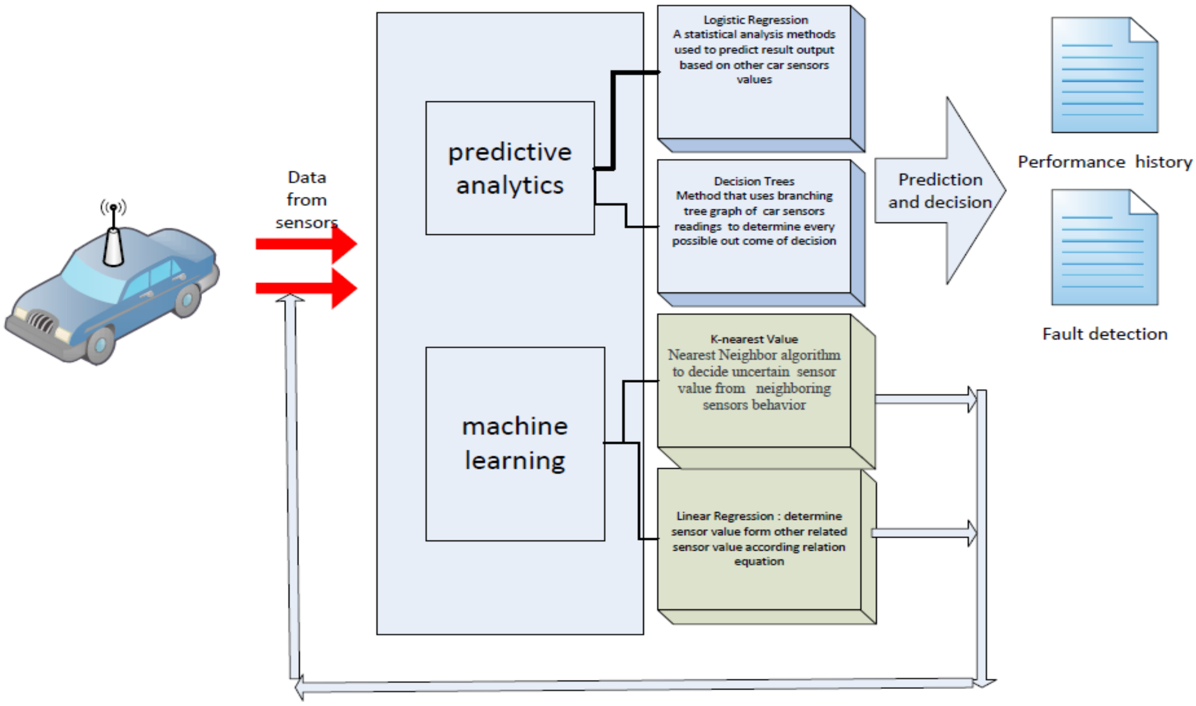


Figure 7: Tow main parts of data analysis in BDMC

6. EXPERIMENTAL REAL APPLICATION

Real application using BDMC has been conducted as a case study. The application is performed on two kinds of test cars, one car is relatively old vehicle 2008 Hyundai- Elentra - spark ignition- Petrol, and the other is new 2018 Hyundai – Tuscan, Compression ignition –Diesel, the sensor data readings from ECU transmitted using Smart-Bus IR Emitter G3, data instantly received in special purpose server to perform as a cloud.

6.1 Test number one:

We have defined sensors that are concerns engine performance , these sensors have a relationship that they check the engine performance in the test drive, the mechanic engineer has define three test drives situations straight and un tilted road , up road and down roads , we define three diving status depending on gas pedal pressure:

1. Idlespeed: no pressure on gas pedal when the vehicle going straight or down
2. Partload: Little pressure on gas pedal to keep speed accelerating
3. Fullload: press on gas pedal to make vehicle accelerate continuously

Accordingly we have defined the following sensors in Table 2 that are related to engine performance and fault detection :

Table 2: Spark injection petrol vehicle, related sensors affects the engine

Sensor name	Sensor check / Unit
Ignition switch	On/ Off to check the ignition electrical spark is ready
Engine speed	the crankshaft spin rotation speed , rotation per hour
Ignition timing cylinder #1:	Primary current connect/disconnect signal is sent to the power transistor to control ignition timing. Ignition is done in sequence of cylinders 1, 3, 4, 2. No unit specified
Ignition timing cylinder #2	
Ignition timing cylinder #3	
Ignition timing cylinder #4	
Injection duration cylinder 1	is the period of time during which fuel enters the combustion chamber from the injector , Unit mili second
Injection duration cylinder 2	
Injection duration cylinder 3	
Injection duration cylinder 4	
Water temperature	Check the the engine cooling water temperature ,Unit C

Data is received by BDMC and directly stands for analysis. Figure 7 depicts a visualization of related sensors behavior, the transmitted data is analyzed and normalized according to the following proceses

1- The received engine rotation speed is normally in rotate per hour , which produced large number of rotation make it hard to make normalization with other sensor readings so the system recalibrates the reading into rotation by minute instead. On the flowing equation

$$Y_R = X_R / t \quad t = 60 \text{ Minutes}$$

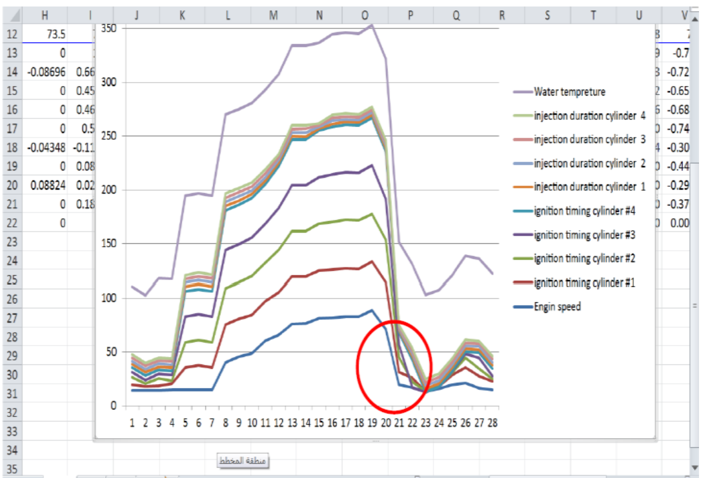


Figure 8: Engine related sensors readings

2- Produce logical regression percentage curve from the received data Figure 9 , the data reading sensors are related and one can affects others so the curve will be changed according to the driving modes mentioned above (Ideal , Part load, Full load) , consider that all of the related sensors reading must be consistent , if there is no fault all the percentage of the sensors will be steady in going up or down with almost same percentage. For each sensor reading we calculate the percentage of change in relation with other sensors ,

$$P = (X_i / X_{i-1}) - 1$$

In no-fault cases, the result of the percentage will be consistent with the logical regression percentage curve , but when percentage goes up or goes down.The fault can be detected in visualized data representation and from percentage variation , as in Figure 10 it depicts default the appear in red circle in Figure 8, we notice that ignition timing cylinders 2, 3 , in time interval 20 to 23 , are not showing reading consistent with the sample regression form , and intersecting other sensors readings on the graph , in calculation there values are not consistent only but contradict the other related sensors changing percentage

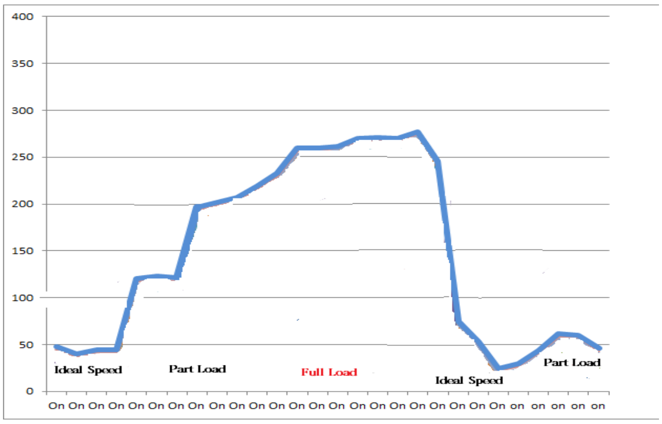


Figure 9: Logical Regression curve mode of related sensors

Actually in this case the fault is not sever and may not affect the vehicle badly , the vehicle will still running , but the engine performance is not fully effective , and this fault might be get worse by the time , so the BDMC will send normal message for the vehicle driver that the vehicle need some check to have better performance , and keep a report and history of this report for repairing and enhancement the vehicle performance.

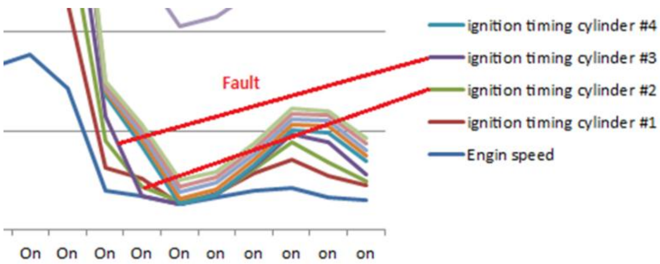


Figure 10: Fault detection in sensors readings

6.2 Test number two

In this test we are going to check faults concerning engine cooling system, this test will be committed by two phases. The first phase considers the sensor for accelerator pedal , and drive the car ups and downs depending on accelerating pedal.:

Table 3: Case 2 diesel injection vehicle , related sensors affects the engine

Sensor name	Sensor check / Unit
Engine speed	Engine speed the crankshaft spin rotation speed , rotation per hour
Cooling fan	Uses electric cooling fans to help pull air through the radiator so that it can keep the engine cool, ON/OFF
Engine coolant temp	Check the engine temperature, and send for fan to work , Range up to 80 C then start cooling fan to work , Unit Celsius
MAP sensor	Manifold Absolute Pressure sensor, used by fuel injection engine determines the engine's air mass flow rate. kPa which is Kilo Pascal a pressure unit.
Engine speed	Engine speed the crankshaft spin rotation speed , rotation per Minute
MAF sensor	Mass Air Flow sensor , determine the engine's air mass flow rateThe engine management system calculates the optimum fuel quantity to be injected from these values. On diesel vehicles the MAF sensor is also used to control exhaust gas recirculation. Gram /Second g/s
Accelerator Pedal Position	This sensor transmits the acceleration pedal position , unit it percentage from 10% up to 60% according vehicle gear.

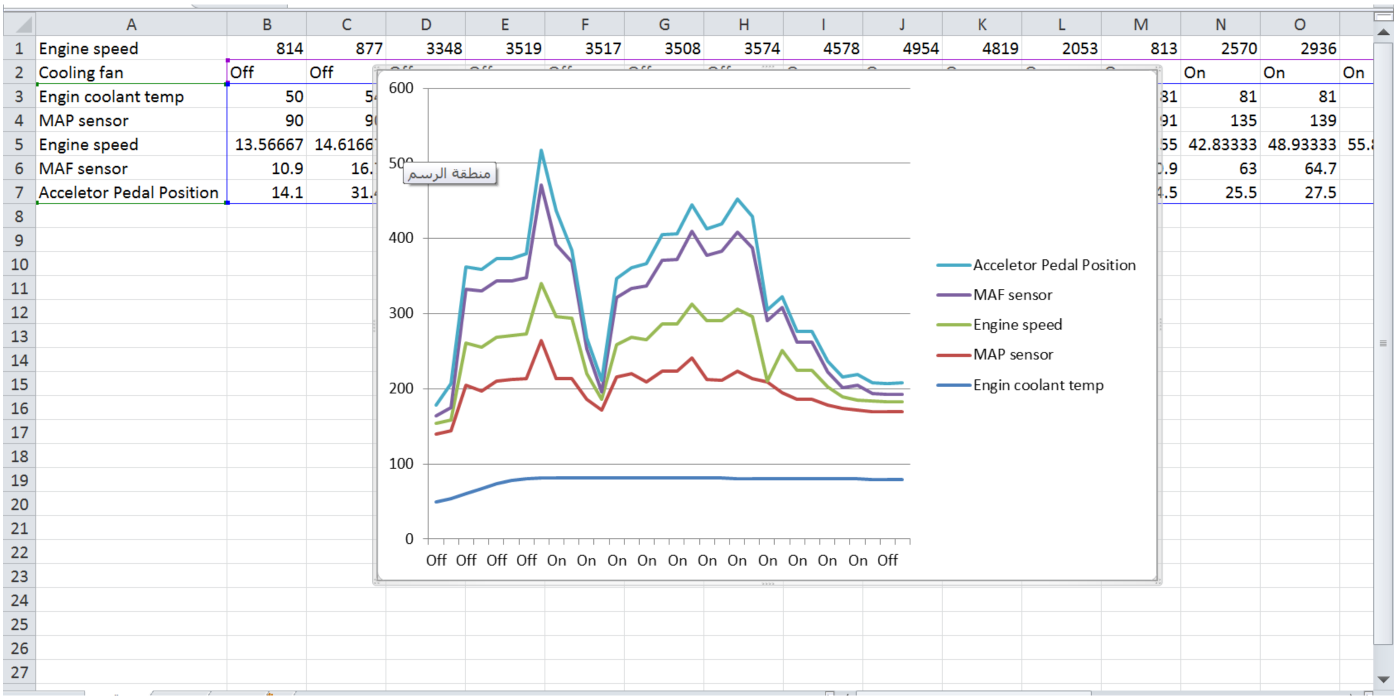


Figure 11: Big Data shows that no faults readings of related sensors regarding cooling engine

Sensors transmit data into BDMC , Figure 11 the reading of sensors are going good , this phase notice the cooling fan is normal , it runs (On) when engine temperature is going above 80 C , when engine cooled down fan will go Off . We discussed in the first test the rest of related sensors defined in table 3 are sensors showing normal, with no faults detected based on the, percentage regression curve, which has no intersection or contradictions in sensors readings.

The second phase of this test is having a sever fault while driving, we have done the same check on sensors shown above in table 3, then we send signal to disconnect the cooling fan while driving. this probability may happens in a real trip, Figure 12 shows a sever fault detection

has occurs, the engine temperature starts to rise above the normal range, at the same time the cooling fan did not work to cool up the engine, in this case the fault is very severe, and if the temperature keeps rising up and the vehicle still running , it will be severely damaged, and it may reach into general failure in engine within very short period of time .

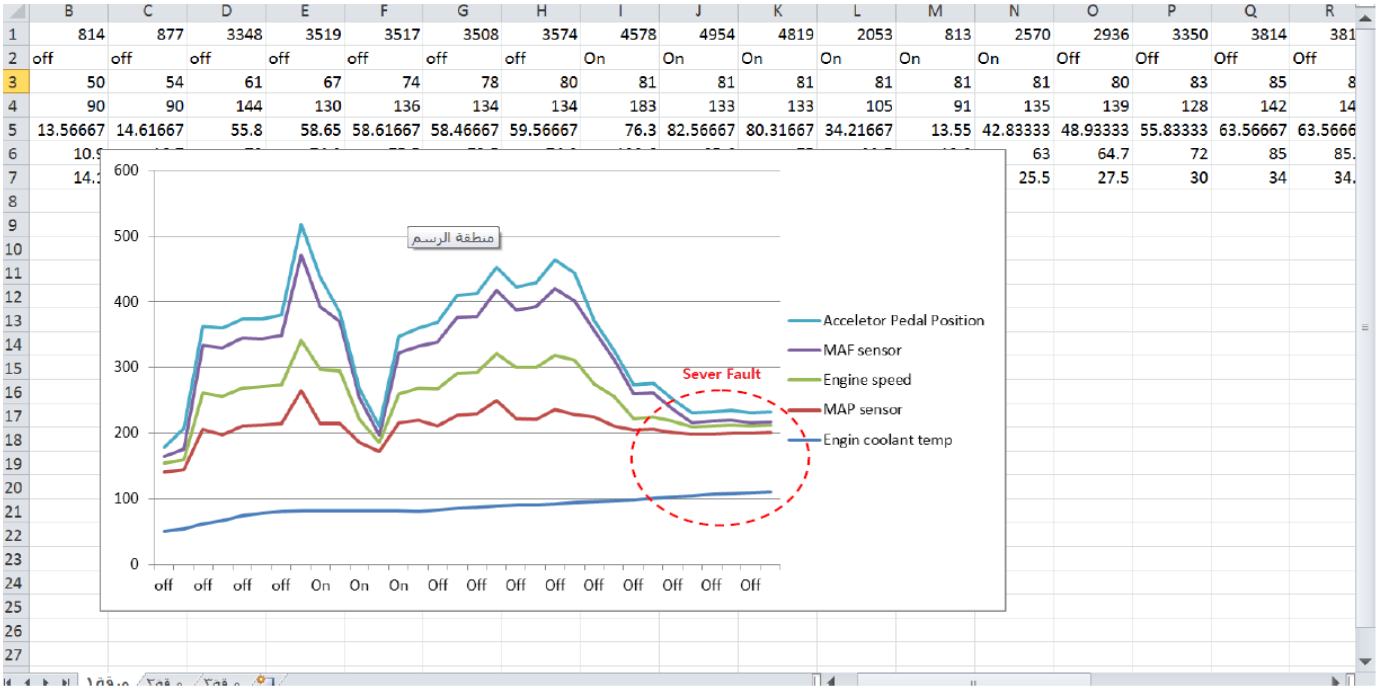


Figure 12: Sever fault detection may lead into general familiar of engine

In this case BDMC will detect this case as sever fault that could damage the engine, the main fault cause is that temperature sensor above normal and get out of range , as soon as the temperature goes up 90 C the BDMC will send urgent message for vehicle driver , urging him/her to stop the vehicle at once , and keep the driver connected with center for further follow- ups and may need to vehicle towing service.

According to mechanical engineers some sensors are categorizes critical sensors , any out of range readings of these sensors cannot be tolerated , sensors like cooling engine ,lubricate oil level , transmission speed , crankshaft position ...etc, these sensors readings should be in the range or else a severe damage on vehicle will occur,. Notice in Figure 12 too, the other sensors starts to get faults readings specially MAP and Engine Speed are getting out of normalized form according to percentage sampling logical regression curve that discussed in the first case

7. RESULTS AND CONCLUSIONS

IoT is defined as connecting Smart devices and machines through wireless to cloud repository, huge amount data is stored in cloud into data primitive forms, this raw data is accumulated in the cloud. Distributed file system Hadoop is used to store unstructured or simi-structured data in data blocks and nodes, smart machines and devices transmit data though wireless gateways to the cloud , BDMC is the main part of this system which provides efficient analysis of big data, including two main parts , fault detection , and machine learning . The first is concerning fault detection, that detect data transmitted from related sensors in machine, as soon as data arrives into the system, fault detection algorithms is applied . Basically two kinds of faults we are dealing with , fault that doesn’t not affect the machine functionality, but reduce the general performance, this fault can be detected from normalized logical regression percentage curve of sensors reading , the second fault is identified as sever, which affect the machine functionality and may reach to general failure, and this can be detected by decision tree. This sever fault normally happened when one vital sensor reading is going out of range, in this case we can avoid much serious problem can lead into general failure and vehicle part destruction . The second part of BDMC is machine learning and lost sensor reading value , this part is depending on one or more other sensor we apply linear regression to anticipate the noisy or lost value , the other option is lost value of independent sensor so we apply KNN algorithm from the nearest neighbors of the same sensor readings. This article presents a case study as prototype of this system is applied on two kinds of vehicle in real trip sensors reading , real readings have been transmitted from ECU of the vehicles though wireless 3G smart control bus to cloud prototype, the results of the two tests vehicle are listed above and in supplementary files ,

REFERENCES

[1] SalaheddinJuneidi New Computing Paradigm: Agent Orientated Engineering and Programming, International Review on Computers and Software 9(6), June 2014

[2] S.Juneidi, G.Vouros ,Survey and evaluation of agent-oriented software engineering main approaches-International Journal of Modelling and Simulation, 2015

[3] Gubbi, J., et al., Internet of Things (IoT): A vision, architectural elements, and future directions. Future Generation Computer Systems, 2013. 29(7): p. 1645-1660.

[4] Madakam, S., R. Ramaswamy, and S. Tripathi, Internet of Things (IoT): A literature review. Journal of Computer and Communications, 2015. 3(05): p. 164. .

[5] Evangelos A, K., T. Nikolaos D, and B. Anthony C, Integrating RFIDs and smart objects into a Unified Internet of Things architecture. Advances in Internet of Things, 2016. 2011.

[6] How Cloud Computing Will Change the Aviation Maintenance Operation, paper white, ATP, © Copyright 2012-2014, ATP., registered in the U.S. and other countries.

[7] Pavithra, D. and R. Balakrishnan. IoT based monitoring and control system for home automation. in Communication Technologies (GCCT), 2015 Global Conference on. 2015. IEEE.

[8] Chun-Wei Tsai, Chin-Feng Lai, Han-Chieh Chao and Athanasios V. Vasilakos Big Data Analytics. Big Data Technologies and Application 2016, Springer, hardcover ISBN:978-3-319-44548-9

[9] Troy Hiltbrand, Behavior-Based Budget Management Using Predictive Analytics, journal Article: *Behavior-Based Budget Management Using Predictive Analytics,DE-AC07-05ID14517, 2015*

[10] 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15) Big Data and Hadoop-A Study in Security PerspectiveB. Saraladevia, N. Pazhanirajaa, P. Victer Paula, M.S. SaleemBashab, P. Dhavachelvanc© 2015 (ISBCC'15)

[11] Automotive Sensors and Electronics Expo 20174th Annual Conference and ExhibitionJune 14–15, 2017 Detroit, Michigan

[12] Festim Halili , Festim Kamberi Performance analysis of classification Algorithms: A case study of Naïve Bayes and J48 in Big DataJournal of Applied Mathematics and Computation (JAMC) 2018, 2(2), 50-57.

[13] Liqiang Pan, Jianzhong Li , K-Nearest Neighbor Based Missing Data Estimation Algorithm in Wireless Sensor Networks, January 2010 ,Wireless Sensor Network 2(02):115-122 DOI: 10.4236/wsn.2010.22016

† Special thanks to Palestine Technical University - Khadoorie for continuous support of research activities