

**PROJECT REPORT**  
**ON**  
**ANALYSIS OF THE**  
**COMMUNICATION NETWORK**  
**OF A BUS**

PROJECT BY: M NISHKAL GUPTA

SHANMUKHA RAJU

GROUP NAME-(DYNAMIC DUO'S 007)

# **INDEX**

- **AIM/GOAL**
- **USED MATERIAL**
- **METHADODOLOGY AND ANALYSIS**
- **FLOW CHART**
- **CONCLUSION**
- **REFERENCES**

## **ACKNOWLEDGEMENT**

Special thanks to Siamak Khatibi for guiding us and providing us with the data required for analyzing and testing our problem.

## **AIM/GOAL**

The main motive of this project is to study the characteristics of the communication network of a bus and to verify whether or not the bus is communicating to the reporting station at equal intervals of time during its journey from source to destination and if it is not reporting at equal intervals then we analyze if it is a system error or random error that is causing unequal intervals of reporting in the communication network. We use “MATLAB” to analyze the problem.

## **MATERIALS USED**

- **btBusdata**(data which contains information about the bus, for example, the bus Unit Id, Trip Number, Line Number, Next stop, Last reporting time etc.) in CSV(comma separated variable) file.
- **Matlab and its Tools and Functions**

## **Note**

System errors are errors that are caused due to a specific parameter in the system for example in the communication systems the system error may be caused due to a defective antenna.

Moreover, system errors have a specific pattern and the values of the error do not change if the same experiment is repeated ‘n’ number of times where as random errors do not follow a pattern and their value changes if the same experiment is repeated ‘n’ number of time.

## METHODOLOGY AND ANALYSIS

Our main motive is to observe the data provided and draw analysis from it. In the data we are given a group of busses which travels between 2 places, place A and place B. During the journey it has to report its information to the reporting station at equal intervals, so our first step is to check whether the bus is reporting at equal intervals or not. If yes then there is no delay in reporting, if not then we have to make a few statistical analysis and find out the reason behind the delay. From the Matlab code, we were successful in finding out the difference between the last reporting times and concluded that the bus wasn't reporting at equal intervals, now our task is to find out the reason behind unequal reporting intervals and study its characteristics.

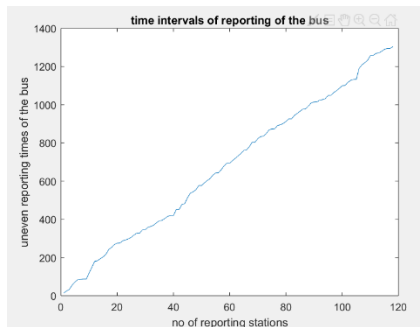


Figure-1

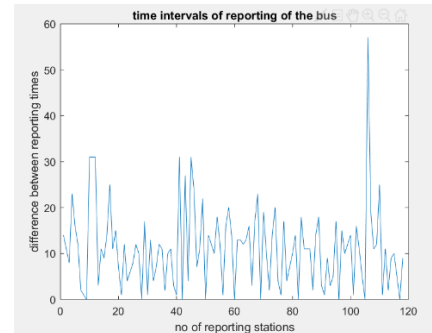


Figure-2

Figure-1 represents the uneven time at which the bus reports (the reporting times are in secs)

Figure-2 represents the difference between uneven times at which bus is reporting (time difference is in secs)

The above graph is for the 'btBusdata.noHeader.csv.000' data for the bus having Unit Id-101222, trip number-41 and line number-6. To study the reason behind the unequal intervals of reporting we must first define equal intervals where the bus has to report (expected reporting time) which can be done by small mathematical manipulation i.e. (total time taken for the bus to reach to the destination / total number of reporting points). Next, we find the difference between expected reporting time and the time where the bus is reporting and plot it.

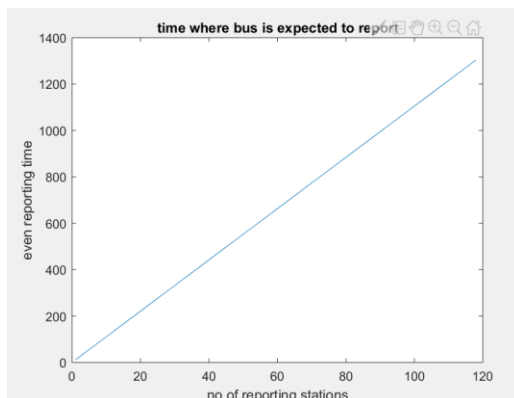


Figure-3

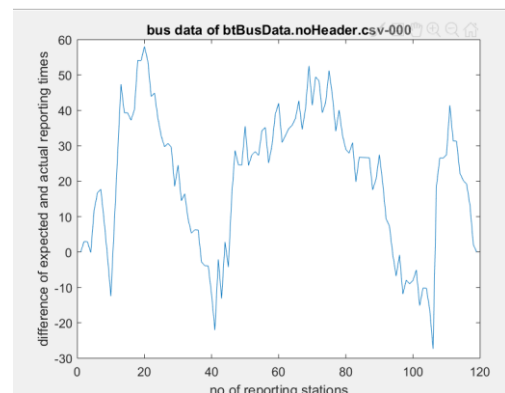


Figure-4

Figure-3 represents equal reporting of bus during its journey (the reporting times are in secs), the total number of reporting stations are 119 for the above data.

Figure-4 represents the difference between actual reporting and expected reporting of the bus (time difference is in seconds) given in data set btBusData.noHeader.csv-000.

From figure 4 we can analyze that few of the differences between the actual and expected reporting are negative, this indicates that at few places the bus is reporting faster than expected time and at other places it's positive which means it's reporting slower than expected time(delay). By looking at the plot we can see that there is a pattern involved with the data i.e. the difference in time reporting at different stations are not randomly caused but they are caused due to systematic errors to be precise it is a scale factor systematic error that is the change in value of the time delay is greater/lesser than the expected value. Moreover, if the experiment is repeated many numbers of times the same error values are obtained which only happens if there is a systematic error because in case of random errors each time you repeat the experiment we obtain new values which are not similar to the previous one. In order to confirm it we have to make hypothesis testing and also verify if the pattern is similar for the same bus taking different trips (each trip may have a different route) to the destination, the data for which is given in btBusData.noHeader.csv-001,002,003 etc.

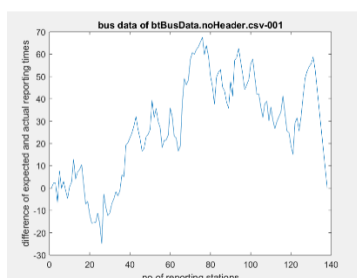


Figure-5

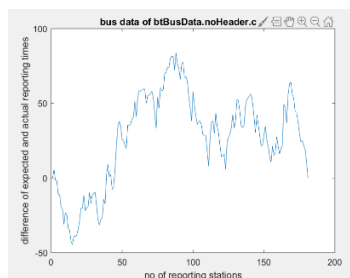


Figure-6

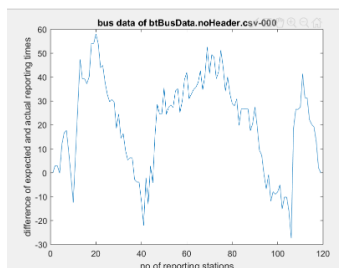


Figure-7

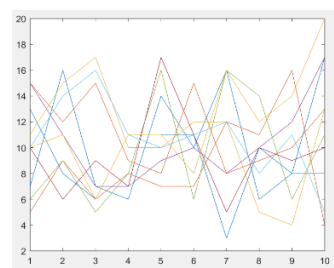


Figure-8

The above 3 pictures (Figure-5,6,7) represent the difference between expected and actual reporting times of the same bus for three different trips (the trips may have different routes to reach the destination). We can see that they all have a pattern and are not random. For reference we can see how a randomly generated data would look from Figure-8.

Next, we perform Kolmogorov-Smirnov hypothesis test on the uneven reporting data of the bus for different trips, the data for which is given in btBusData.noHeader.csv-001,002,003 etc. By performing the test, we concluded that the uneven reporting data of the bus for different trips belong to the same continuous distribution, by this we can conclude that the delay of bus reporting from expected time is caused due to a common factor. Moreover, this makes our statement strong that the delay is caused due to systematic error.

NOTE- kstest2(x,y) returns a test decision for null hypothesis that the data in vector x and y belong to the same continuous function if the output is logical 0 and vice-versa.

### **ANALYSIS OF DIFFERENCES BETWEEN EXPECTED AND ACTUAL REPORTING TIME OF A BUS FOR A SINGLE TRIP**

In order to study the characteristics of the data it is always recommended to observe the histogram and then draw theoretical calculations from it. Below is the histogram using Scott's formula( $h=3.49*\text{std}(\text{data})/\text{length}(\text{data})^{(1/3)}$ ) for the data of the differences between expected and actual reporting time of a bus obtained from the data set "btBusData000". By viewing the histogram, we observe that it is normally distributed but when performed hypothesis on the data it was rejected at 5% significance level. I used Chi-square goodness of fit test to perform hypothesis.

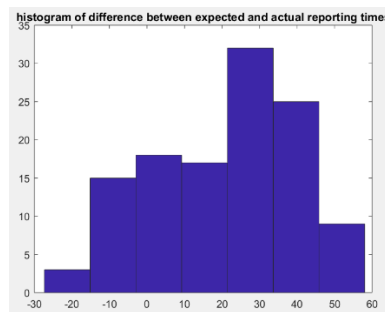


Figure-7

Moreover, the correlation between the time intervals(delays) is found to be very high (0.8807) which tells us that they are highly dependent on each other. We also calculated the mean between consecutive data points and observed that the mean is not constant but differs with changes. Hence by this we can draw the observation that the delay is neither independent nor identical i.e. it is not I.I.D (independent and identical distribution). Hence all these analyses together gives us information that the delay is caused due to a parameter which it is depending on such as an antenna etc. which leads to systematic error.

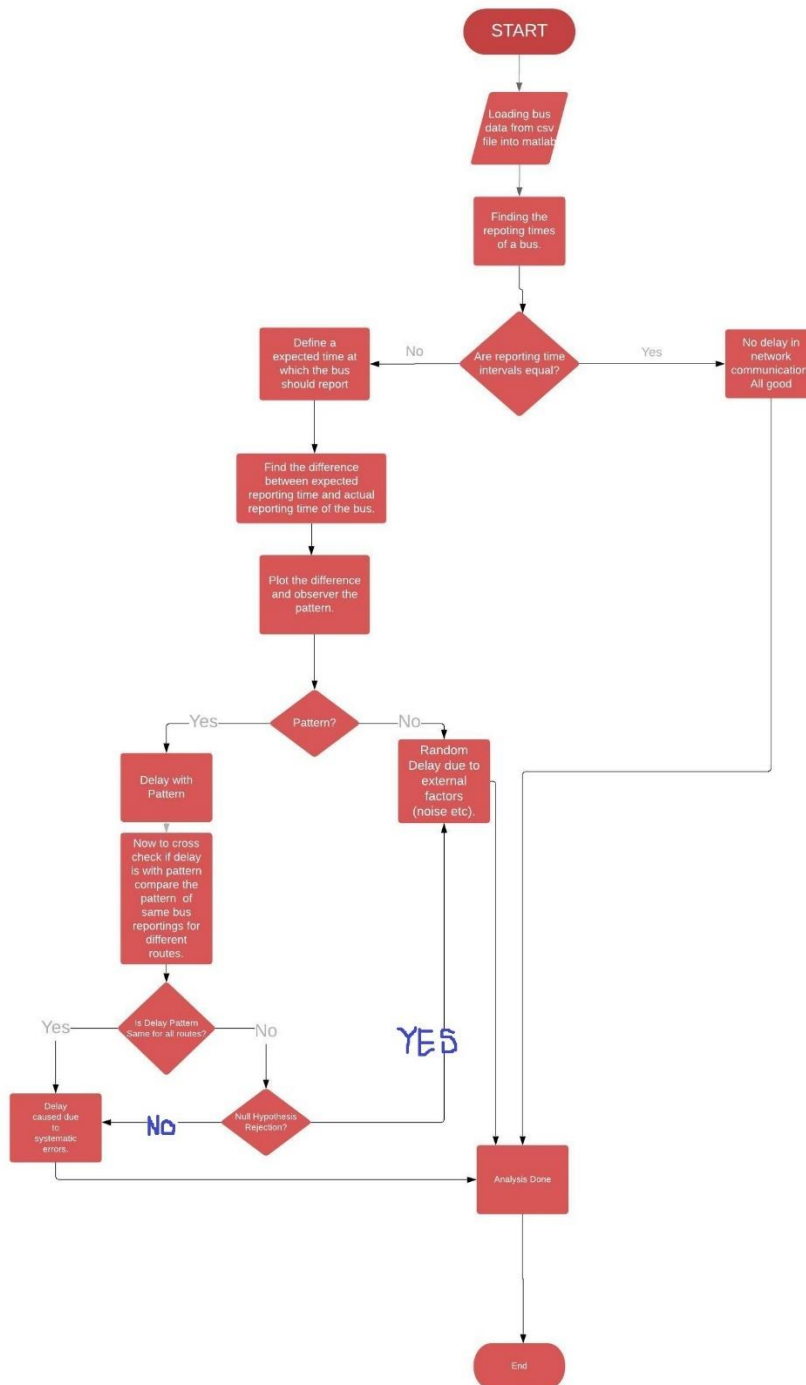
### **NOTE**

We perform analysis on a bus taking different trips as well as a bus taking a single trip. Analysis on bus taking different trips is done to verify whether the pattern of delay is same for all the trips or not, if it's same then we can assume that the delay is caused due to systematic error and not a random error and then go for a bus taking single trip and analyze the characteristics of the

systematic error in the network by performing statistical operations such as hypothesis, histogram plotting, finding mean etc. which are mentioned in the report above.

## FLOW CHART

Below is the flow chart which gives us a brief overview.



## **CONCLUSION**

We can conclude by above analysis that the bus is not reporting at equal intervals to the reporting station and the reason for not reporting at equal intervals is a delay that is caused due to systematic error which we have analyzed using appropriate statistical techniques mentioned above in the report.

## **REFERENCES**

- <https://in.mathworks.com/help/stats/chi2gof.html#btv1kvu-1>
- <https://www.physics.umd.edu/courses/Phys276/Hill/Information/Notes/ErrorAnalysis.html>
- <https://in.mathworks.com/help/matlab/ref/corrcoef.html>
- **Probability and Statistics by S.Chand**
- [https://en.wikipedia.org/wiki/Independent\\_and\\_identically\\_distributed\\_random\\_variables](https://en.wikipedia.org/wiki/Independent_and_identically_distributed_random_variables)
- **Lecture PPT's provided by Siamak Khatibi**

