pseudo code: cleaning data by descending order in 4\_point method

Inputs:

all\_points\_1: Traffic data read from database for the 4 consecutive sensors

all\_points\_2: Data frame used to save the outputs

seq\_list: Sequence of each sensor

tt\_threshold: Travel time threshold pre-determined by the user to eliminate the abnormally long trip of the two end segments

Outputs: all\_points\_2

Get the minimum sensor sequence from the seq\_list, save as seq\_2

Get the maximum sensor sequence from the seq\_list, save as seq\_1

Get the unique mac\_address from the all\_points\_1, save as mac\_list

FOR each mac\_address

Get the subset of the traffic data of this mac\_address, save as subset\_data

IF length of the sequence for the subset\_data==4

Get the first index of seq\_1, save as temp\_ind

Copy the row in the subset\_data with the index temp\_ind, save as one\_row

Add one\_row to the subset\_data

Get the indices of seq\_1, save as index\_1

FOR each index i in index\_1 except the last one

IF index\_1[i+1]-index\_1[i]>=3

Get the subset data from subset\_data[index\_1[i] : (index\_1[i+1]-1)], save as df\_sub

IF length(unique(df\_sub$seq))==4

Find the first index of index\_2, save as first\_ind

Subset the data df\_sub[1: first\_ind], save as df\_sub

IF all difference of the adjacent indices <= 0 & unique sequence is equal to 4

Get the subset data of df\_sub from the first row to the second to the last one, save as df\_1

Rename the columns of df\_1 as c("ip\_o","time\_o","location\_o","mac\_o","seq\_o")

Get the subset data of df\_sub from the second row to the last one, save as df\_2

Rename the columns of df\_2 as c("ip\_d","time\_d","location\_d","mac\_d","seq\_d")

Form the OD data by column combind the df\_1 and df\_2, save as df\_3

Calculate the trip time by the difference of timestamp in column time\_o and time\_d,

save to column ttime

IF all ttime <= tt\_threshold

Save the data to the output data frame all\_points\_2

END IF

END IF

END IF

END IF

END FOR

END IF

END FOR