Trace Route Packet Analysis, INFSCI 2310

Taiwo Omileye

12/10/2021

## AIM:

In this project, the traceroute event of an end-to-end path was examined through data analysis of 5 different domain names/IP address from my residence to the destination path on daily basis. This is to examine the stability of the time-delay and number of routers require as the datagram passes through the end-to-end path. Three measurements on different hours were taken beginning on 11/19/2021 consecutively per day which represents the 15 measurements as x variables (5 domains multiply 3) according to the requirements for a consecutive number of 7 days which also represents y variables according to the requirements. Therefore, the total data points generated is 105 observations (O), x\*y datapoints (15 X 7) as shown below. These measurements were taken in every three-hour interval daily. My residence is located around Mcknight road. The five domains I used for this analysis project which is analogous to the corresponding IP address include www.pitt.edu, www.google.com, www.cmu.edu, www.yahoo.com, and www.amazon.com respectively.

## MATERIAL USED:

R programming language and required packages, traceroute software tool, excel notebook in csv format.

Figure A.

## METHOD:

Figure A above describe the virtual path my packet takes as I route it from my home to the 5 chosen destination where each router or hops represent the network layer along the path which their total sum for a domain is represented by the MaxHops in the observations (O). The time-delay (milli seconds, ms) at every hop/router for a given destination is represented by HopsTime O. while the total time-delay for all routers in a given domain is Time O. The specific day out of 7 days is represented by day variable O, and the type of measurement out of the three estimation for a day is represented by Measurement variable in O. The granularity used to examined stability are Time (ms) & MaxHops number (integer) in Observation O. The reason for these metrics are varying value of observational effects for each measurement over the 7 days. The stability of the path is examined through the different plot below and the summary page of each variable. The findings are also summarized in the conclusion part.

## OBSERVATION:

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.4 v dplyr 1.0.7  
## v tidyr 1.1.3 v stringr 1.4.0  
## v readr 2.0.1 v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

## corrplot 0.90 loaded

## Loading required package: lattice

##   
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':  
##   
## lift

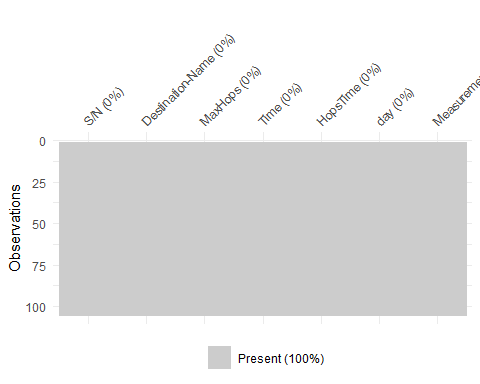
## Rows: 105 Columns: 7

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (2): Destination-Name, HopsTime  
## dbl (5): S/N, MaxHops, Time, day, Measurement

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## Rows: 105  
## Columns: 7  
## $ `S/N` <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, ~  
## $ `Destination-Name` <chr> "www.pitt.edu", "www.google.com", "www.cmu.edu", "w~  
## $ MaxHops <dbl> 2, 7, 3, 8, 3, 3, 8, 3, 7, 2, 2, 7, 3, 12, 3, 5, 2,~  
## $ Time <dbl> 9, 53, 15, 234, 18, 17, 234, 13, 47, 9, 11, 49, 19,~  
## $ HopsTime <chr> "[2,7]", "[1,9,9,7,6,6,15]", "[2,2,11]", "[1,9,34,3~  
## $ day <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, ~  
## $ Measurement <dbl> 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 1, 1, ~

### The above indicates that domain ‘www.pitt.edu’ is the closest Autonomous system, AS, to my dormitory as it has lowest routers/MaxHops, follow by ‘www.cmu.edu’,‘www.amazon.com’,‘www.google.com’ and ‘www.yahoo.com’. This implies domain ‘www.yahoo.com’ is the longest Autonomous system, AS. to my residence.



## S/N Destination-Name MaxHops Time   
## 105 5 8 57   
## HopsTime day Measurement   
## 86 7 3

## # A tibble: 8 x 2  
## MaxHops n  
## <dbl> <int>  
## 1 1 3  
## 2 2 18  
## 3 3 38  
## 4 5 8  
## 5 7 12  
## 6 8 15  
## 7 11 1  
## 8 12 10

### As it can be seen above, the dominant Hops/routers for a given path among the chosen domains is 3 routers indicating out of 5 domains/ autonomous systems, AS. The dominant Hops for an AS has 38 number of occurences and the fact that median of MaxHops equal 3 in the summary figure below.

## # A tibble: 86 x 2  
## HopsTime n  
## <chr> <int>  
## 1 [0,1,5,4,7,16,16,15] 1  
## 2 [1,1,12] 2  
## 3 [1,1,4,5,4,15,5,4] 1  
## 4 [1,1,5,5,5,13,14,35,37,35,33,35] 1  
## 5 [1,1,6,4,7,8,14,35,35,35,33,35] 1  
## 6 [1,1,6] 1  
## 7 [1,1,9] 6  
## 8 [1,10,34,36,35,39,36,39] 1  
## 9 [1,14,34,36,36,38,35,40] 1  
## 10 [1,17,7,7,5,8,15] 1  
## # ... with 76 more rows

### As shown above, out of 105 observations I got. We have 5 unique domains, 8 unique routers’ number path, 57ms unique total routers’ delay and 7 unique days with 3 measurements for each domain.

## # A tibble: 57 x 2  
## Time n  
## <dbl> <int>  
## 1 1 1  
## 2 2 1  
## 3 3 2  
## 4 7 3  
## 5 8 8  
## 6 9 6  
## 7 10 2  
## 8 11 7  
## 9 12 5  
## 10 13 3  
## # ... with 47 more rows

### The dominants delay for the end-to-end path among my chosen domains is 8 milli-seconds as seen from the above.

## [1] "[2,7]" "[1,9,9,7,6,6,15]"   
## [3] "[2,2,11]" "[1,9,34,35,42,39,35,39]"   
## [5] "[3,2,13]" "[4,1,12]"   
## [7] "[1,14,34,36,36,38,35,40]" "[3,1,9]"   
## [9] "[1,7,7,6,6,5,15]" "[2,7]"   
## [11] "[4,7]" "[1,7,6,9,6,5,15]"   
## [13] "[3,1,15]" "[2,2,7,7,6,8,16,35,45,38,33,35]"   
## [15] "[2,2,9]" "[6,1,5,5,9]"   
## [17] "[2,7]" "[1,7,7,8,6,8,15]"   
## [19] "[4,1,10]" "[2,9,34,38,38,39,36,37]"   
## [21] "[4,11]" "[1,29,5,4,16,16,4,4]"   
## [23] "[4,2,9]" "[2,37,37,35,36,40,36,38]"   
## [25] "[1]" "[2,6]"   
## [27] "[2,1,4,5,4,6,5,5]" "[2,2,9]"   
## [29] "[2,1,5,5,5,6,14,40,37,35,33,35]" "[1]"   
## [31] "[3,7,7,6,5]" "[1,1,4,5,4,15,5,4]"   
## [33] "[1,1,9]" "[1,1,6,4,7,8,14,35,35,35,33,35]"   
## [35] "[3,2,7,6,6]" "[14,3,9,15,17,19,35]"   
## [37] "[2,9,34,36,37,40,36,37]" "[1,1,12]"   
## [39] "[2,1,4]" "[1,7]"   
## [41] "[2,7]" "[5,7,11,7,6,5,15]"   
## [43] "[3,1,8]" "[1,10,34,36,35,39,36,39]"   
## [45] "[1,1,9]" "[6,1,4,7,5]"   
## [47] "[1,1,5,5,5,13,14,35,37,35,33,35]" "[1,1,9]"   
## [49] "[2,1,5]" "[1,7]"   
## [51] "[1,6]" "[1,6,6,10,5,7,15]"   
## [53] "[3,3,28]" "[27,1,5,5,6,5,15,38,37,36,33,37]"  
## [55] "[3,9,9]" "[1,9,10]"   
## [57] "[2,1,5,4,4,11,14,34,38,36,35,35]" "[1,2,9]"   
## [59] "[4,1,6,6,6,16,4,15]" "[2,7]"   
## [61] "[1,7]" "[0,1,5,4,7,16,16,15]"   
## [63] "[1,1,9]" "[2,9,34,35,35,37,36,35]"   
## [65] "[2,9,9]" "[4,4,6,6,5]"   
## [67] "[3,1,5,8,8,7,13,38,38,37,33,35]" "[1,3,11]"   
## [69] "[1,17,7,7,5,8,15]" "[4,6]"   
## [71] "[3,7]" "[2,2,5]"   
## [73] "[1,1,12]" "[2,1,5,6,5,14,18,35,35,35,33]"   
## [75] "[1,1,9]" "[1,7]"   
## [77] "[1,9,34,35,37,39,35,35]" "[2,5,15]"   
## [79] "[2,2,4]" "[3,8,6,7,8]"   
## [81] "[1,6,6,7,6]" "[1,6,12,15,5,5,15]"   
## [83] "[3,3,9]" "[1,9,35,35,36,39,35,36]"   
## [85] "[2,1,9]" "[2,5,14]"   
## [87] "[3,2,4,5,4,17,14,35,37,35,33,35]" "[3,1,15]"   
## [89] "[2,6,6,6,5,5,15]" "[1,7,7,8,5]"   
## [91] "[5,7]" "[4,7,6,7,5,6,16]"   
## [93] "[2,1,9]" "[1,8,35,37,35,37,35,36]"   
## [95] "[3]" "[1,9,9]"   
## [97] "[1,2,5,4,4,14,13,35,36,35,34,36]" "[1,1,9]"   
## [99] "[1,1,6]" "[1,6]"   
## [101] "[1,7]" "[3,7,8,7,7,5,15]"   
## [103] "[2,5,9]" "[3,1,6,6,8,6,17,36,36,35,33,33]"   
## [105] "[3,9,12]"

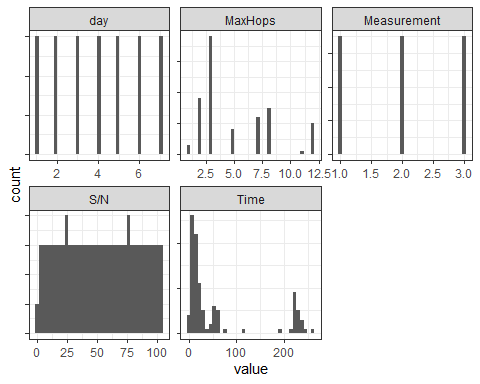
### The time-taken in milli-seconds at each router level along the end-to-end path is also displayed above, for example the first domain (www.pitt.edu) took 2 Hops/router where each delay are 2ms and 7ms respectively and ditto to others.

# Basic Statistics of All Variables

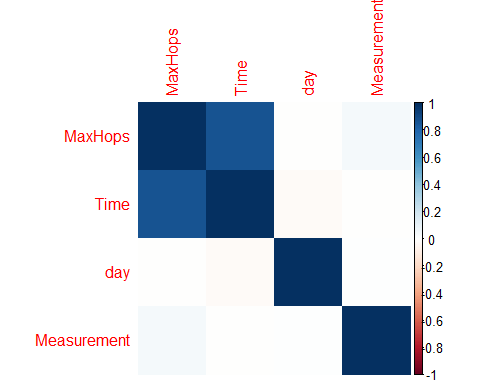
## S/N Destination-Name MaxHops Time   
## Min. : 1 Length:105 Min. : 1.000 Min. : 1.00   
## 1st Qu.: 27 Class :character 1st Qu.: 3.000 1st Qu.: 11.00   
## Median : 53 Mode :character Median : 3.000 Median : 20.00   
## Mean : 53 Mean : 5.029 Mean : 63.48   
## 3rd Qu.: 79 3rd Qu.: 7.000 3rd Qu.: 59.00   
## Max. :105 Max. :12.000 Max. :261.00   
## HopsTime day Measurement  
## Length:105 Min. :1 Min. :1   
## Class :character 1st Qu.:2 1st Qu.:1   
## Mode :character Median :4 Median :2   
## Mean :4 Mean :2   
## 3rd Qu.:6 3rd Qu.:3   
## Max. :7 Max. :3

### Out of my domains/Autonomous system (AS), the minimum total router/MaxHops is 1, its median value is 3 while the maximum one is 12, this is an indication that the lowest AS path to my homen takes 1 router part while the longest AS path has 12 routers along its end-to-end path. The minimum delay for all path is 1ms and the maximum delay is 261ms. The mean for this total delay is 63.48ms while the median is 20, this is an indication of left-skewness among my domains.

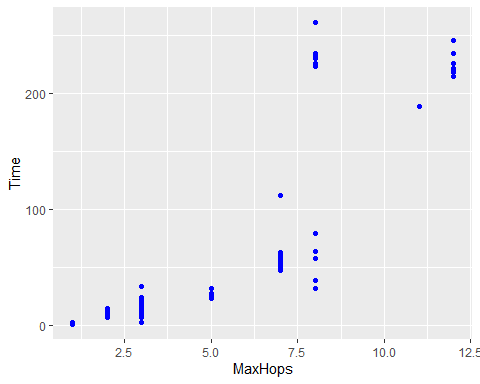
# Distribution of All variables



# Distribution of correlation among variables

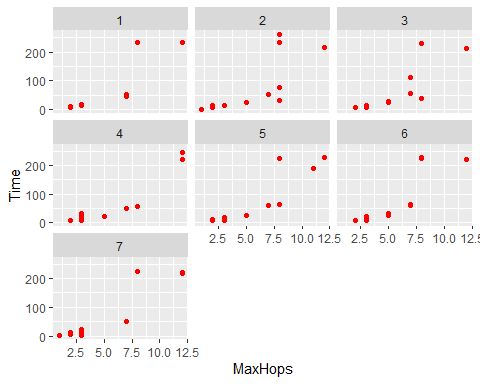


# Scatter Plot between Time and Routers



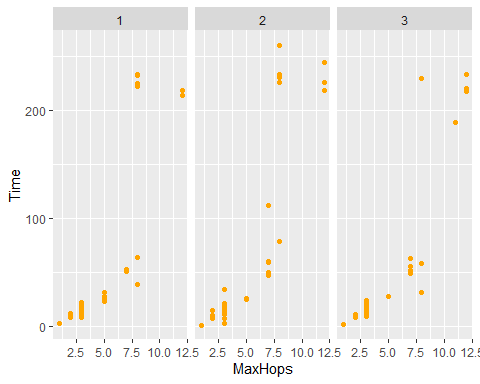
### The above correllation plot with scatter plot also indicate a slight positive correllation (0.8) between the total number of routers/MaxHops and time-taken/delay for a given domain, this explains the fact that the higher the path the datagram take to reach its end, the higher will be the delay

# Scatter Plot between Time and Routers over 7 Days



### The above shows my measurement over the period of 7 days interval at each facet at different combination of total routers delay/Time and total number of routers required/MaxHops for a given domains/destination. Eventhough, the 7 pictures looks similar, however they are slightly different due to different day of their measurement. The shorter the destination path of datagram propagation, the more stable it is delay-router combination. While the longer the destination path of datagram propagation, the more instability is observed in the delay-router combination. This noticeable difference, however can be attributed to different path the datagram take at each day.

# Scatter plot between Time and Routers over 3 Measurments in a Day



### The above also shows my measurement over three-hour interval, even though it looks similar, yet the different can be attributted to different routing path and varying activities of the day among the 5 domains.

# Statistics of All Variables

Data summary

|  |  |
| --- | --- |
| Name | Piped data |
| Number of rows | 105 |
| Number of columns | 7 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 2 |
| numeric | 5 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

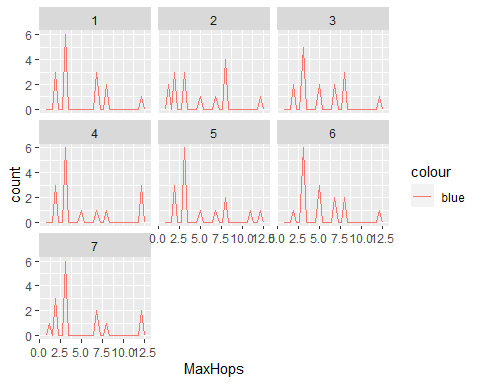
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| Destination-Name | 0 | 1 | 11 | 14 | 0 | 5 | 0 |
| HopsTime | 0 | 1 | 3 | 32 | 0 | 86 | 0 |

**Variable type: numeric**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| S/N | 0 | 1 | 53.00 | 30.45 | 1 | 27 | 53 | 79 | 105 | ▇▇▇▇▇ |
| MaxHops | 0 | 1 | 5.03 | 3.21 | 1 | 3 | 3 | 7 | 12 | ▇▁▂▂▂ |
| Time | 0 | 1 | 63.48 | 83.80 | 1 | 11 | 20 | 59 | 261 | ▇▁▁▁▂ |
| day | 0 | 1 | 4.00 | 2.01 | 1 | 2 | 4 | 6 | 7 | ▇▃▃▃▇ |
| Measurement | 0 | 1 | 2.00 | 0.82 | 1 | 1 | 2 | 3 | 3 | ▇▁▇▁▇ |

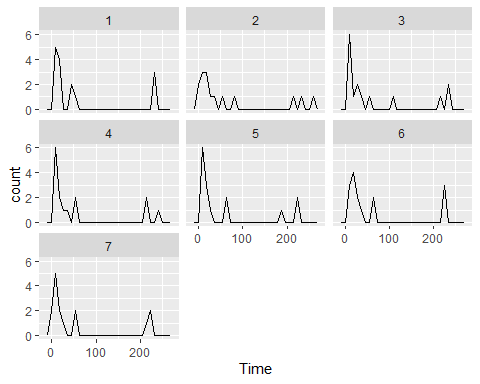
### The mean and standard deviation of the total delay/Time are 63.5ms and 83.8ms which is uneven distribution, similarly for the number of router require for a domain, its mean and standard deviation are 5 and 3 respectively.

# Distribution of the Routers over 7 Days



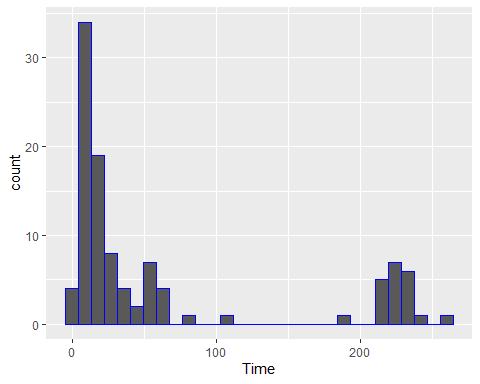
### Above is the distribution of router’s number over the 7-day interval of measurement, the figure becomes stable over a long period of time. For example, figure day 1 and day 7 look perfectly equal while daily picture interval is different.

# Distribution of the Delay over 7 Days



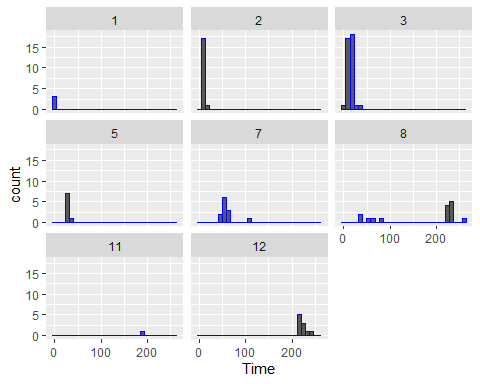
### Similar scenario is observed just like in the MaxHops number over the distribution of total delay over a given domain. Even though, on daily basis each picture looks very different but over a long period of time, the total delay become stabilize. This can be attributted to the stochastic effect of probabilty over a long term.

# Histogram of the Delay for 5 chosen Domains



### Right-skewed distribution of the delay, mean is greater than the median indicating uneven distribution. More delay are tends toward the left side than the middle.

# Distribution of the Delay over Number of Routers



### The distribution of Total delay measurement at each router over the MaxHops require at different domains level. The shortest path domain required 1 router.

## CONCLUSION:

The aftermath measurement of datagram trace routing over the 5 domains (www.pitt.edu“,”www.google.com“,”www.cmu.edu“,”www.yahoo.com“,”www.amazon.com") include:

* The dominant Hops/routers for a given path among the chosen domains is 3 routers indicating out of 5 domains/ autonomous systems, AS. The dominant Hops for an AS has 38 number of occurrences with median value of MaxHops equal 3. This implies majority of my destination require 3 routers.
* Out of the 105 observations I got. We have 5 unique domains/IP destination, 8 unique routers’ number path plus 3 higher than the domain uniqueness, 57ms unique total routers’ delay/Time measurements for each domain. This means time delay changes a lot for all domains over the measurement and the dominants delay for the end-to-end path among my chosen domains is 8 milli-seconds.
* The smallest delay at each router for a domain takes 1ms from the first Hops while longest delay at the last Hops takes 40ms, this can be attributed to the transmission delay and policy requirement over intra-Autonomous system (AS) protocol from the first Hops over the inter-Autonomous system (AS) protocol from another end, last Hops delay.
* There is a positive correlation between the total number of routers/MaxHops for a path and time-taken/delay for a given domain, this explains the fact that the higher the path the datagram take to reach its end, the higher will be the delay. This explain the cause and effect relationship between the two variables.
* Even though, the 7 measurement moving average looks similar between the delay-router combination plot, however they are slightly different due to different day of their measurement. The shorter the destination path of datagram propagation, the more stable it is delay-router combination. While the longer the destination path of datagram propagation, the more instability is observed in the delay-router combination. This noticeable difference, however can be attributed to different path the datagram take at each day due to less or more busy link. Similar effect is also observed for the 3 measurements in a given day.
* The mean and standard deviation of the total delay/Time are 63.5ms and 83.8ms, similarly for the number of router require for a domain, its mean and standard deviation are 5 and 3 respectively which indicates an uneven distribution.
* For this experiment and project, it is observed that the stability of the end-to-end path of datagram for a given domain becomes stable over a long term period but becomes unstable over a short term period due to stochastic effects, correlation between delay & MaxHops among others.