ANALYSING INTERNET CONNECTION TRENDS DURING CHRISTMAS IN UNITED STATES

INTRODUCTION :

I wanted to analyse the Internet connection Trends during social events like festivals.

I chose christmas as it is one of the major festivals and public holidays in United States. Through, my report I try to examine the changes in internet connections during Christmas holiday season.

I am trying to analyse human group behaviour during the holiday season from these changes in Internet connections i.e. preferance of people for being out in the real world during holidays vs. Being connected to Internet . As we know ,load on servers due to high network traffic lead to bad connections.Thus depending on peoples choices,traffic will increase or decrease and network connection statistics will vary.

While this study has only been done on one holiday, we can further conduct this on some major holidays like 4^{th} of July, Thanksgiving, New Years, Black Friday to further characterize and better understand human dynamics and find a relation if any

METHODOLOGY :

In this report I examine the some of the basic network connectivity measures of different connections. The network connectivity measures used are :
1.Average RTT.
2.Download Throughput.
3.Percentage of Packets Retransmitted.

I have used median as aggregating measure since it is one of the least affected measures by big variations in data amongst all aggregating functions.It is hence one of the best representatives for the most occuring value for that function.

The data for connections was collected using Measurement Lab Dataset on Google Big Query over the years 2011 and 2012 as we want to establish a comparison over the most recent trends.

While querying,

results from only NDT data were considered(project=0) as they were considered to be more specific and accurate for network diagnostics done.

Results from only server to client tests were considered as wanted to diagnose client-oriented values for network diagnostic tests
(connection_spec.data_direction = 1)

Results were for only those data were considered when tests had been completed. As if not,then the test results cannot correctly estimate the connection performance
(

```
web100_log_entry.is_last_entry = True

web100_log_entry.snap.HCThruOctetsAcked >= 8192

(web100_log_entry.snap.SndLimTimeRwin +
 web100_log_entry.snap.SndLimTimeCwnd +
 web100_log_entry.snap.SndLimTimeSnd) >= 9000000
```

```
(web100_log_entry.snap.SndLimTimeRwin +
web100_log_entry.snap.SndLimTimeCwnd +
web100_log_entry.snap.SndLimTimeSnd) < 3600000000

(web100_log_entry.snap.State == 1
        OR (web100_log_entry.snap.State >= 5
            AND web100_log_entry.snap.State <= 11))

)
```

Queries used to extract and process the data were as follows :

Average RTT of connections from the country :

The following query extracts the median of Average RTT values of the country(here United States) over a period of one week(here Nov 15 – Nov 21)

The condition to get Average RTT values is :
web100_log_entry.snap.SumRTT/web100_log_entry.snap.CountRTT
We have further averaged it over IPs before calculating the median for the country as a whole.

Here it makes sense to exclude results of tests with fewer than 10 round trip time samples, because there are not enough samples to accurately estimate Average RTT
(web100_log_entry.snap.CountRTT > 10)

Final Query :

```
SELECT percentile_cont(0.5) OVER (ORDER BY rtt)
FROM (
SELECT
web100_log_entry.connection_spec.remote_ip AS ips,
AVG(web100_log_entry.snap.SumRTT/web100_log_entry.snap.CountRTT) AS rtt
FROM [measurement-lab:m_lab.2011_11]
WHERE
IS_EXPLICITLY_DEFINED(web100_log_entry.connection_spec.remote_ip)
AND IS_EXPLICITLY_DEFINED(connection_spec.client_geolocation.country_name)
AND connection_spec.client_geolocation.country_name='United States'
AND IS_EXPLICITLY_DEFINED(web100_log_entry.log_time)
AND web100_log_entry.log_time > PARSE_UTC_USEC('2011-11-15 00:00:00') / POW(10,
6)
AND web100_log_entry.log_time < PARSE_UTC_USEC('2011-11-21 23:59:59') / POW(10,
6)
AND IS_EXPLICITLY_DEFINED(web100_log_entry.connection_spec.local_ip)
AND IS_EXPLICITLY_DEFINED(web100_log_entry.snap.HCThruOctetsAcked)
AND IS_EXPLICITLY_DEFINED(web100_log_entry.snap.SndLimTimeRwin)
AND IS_EXPLICITLY_DEFINED(web100_log_entry.snap.SndLimTimeCwnd)
AND IS_EXPLICITLY_DEFINED(web100_log_entry.snap.SndLimTimeSnd)
AND IS_EXPLICITLY_DEFINED(project)
AND project = 0
AND IS_EXPLICITLY_DEFINED(connection_spec.data_direction)
AND connection_spec.data_direction = 1
AND IS_EXPLICITLY_DEFINED(web100_log_entry.is_last_entry)
AND web100_log_entry.is_last_entry = True
AND web100_log_entry.snap.HCThruOctetsAcked >= 8192
AND (web100_log_entry.snap.SndLimTimeRwin +
     web100_log_entry.snap.SndLimTimeCwnd +
```

```
     web100_log_entry.snap.SndLimTimeSnd) >= 9000000
AND (web100_log_entry.snap.SndLimTimeRwin +
     web100_log_entry.snap.SndLimTimeCwnd +
     web100_log_entry.snap.SndLimTimeSnd) < 3600000000
AND IS_EXPLICITLY_DEFINED(web100_log_entry.snap.MinRTT)
AND IS_EXPLICITLY_DEFINED(web100_log_entry.snap.SumRTT)
AND IS_EXPLICITLY_DEFINED(web100_log_entry.snap.CountRTT)
AND web100_log_entry.snap.CountRTT > 10
AND (web100_log_entry.snap.State == 1
        OR (web100_log_entry.snap.State >= 5
            AND web100_log_entry.snap.State <= 11))
GROUP BY ips )
```

Download Throughput of connections from the country :

The following query extracts the median of Download Throughput values of the country(here United States) over a period of one week(here Nov 15 – Nov 21)

The condition to get Download Throughput values is :
web100_log_entry.snap.HCThruOctetsAcked/(web100_log_entry.snap.SndLimTimeRwin + web100_log_entry.snap.SndLimTimeCwnd + web100_log_entry.snap.SndLimTimeSnd)
We have further averaged it over IPs before calculating the median for the country as a whole.

Also we havent considered those NDT tests where
web100_log_entry.snap.CongSignals = 0
i.e. The test ends during slow start and never reaches congestion. As this happens if the test was interrupted by the user or due to some errors.

Final Query :

```
SELECT percentile_cont(0.5) OVER (ORDER BY thru)
FROM (SELECT web100_log_entry.connection_spec.remote_ip AS ips ,
AVG(web100_log_entry.snap.HCThruOctetsAcked/
(web100_log_entry.snap.SndLimTimeRwin + web100_log_entry.snap.SndLimTimeCwnd +
web100_log_entry.snap.SndLimTimeSnd)) AS thru,
FROM [measurement-lab:m_lab.2011_11] WHERE
   IS_EXPLICITLY_DEFINED(web100_log_entry.connection_spec.remote_ip)
   AND IS_EXPLICITLY_DEFINED(connection_spec.client_geolocation.country_name)
   AND connection_spec.client_geolocation.country_name='United States'
   AND IS_EXPLICITLY_DEFINED(web100_log_entry.log_time)
   AND web100_log_entry.log_time > PARSE_UTC_USEC('2011-11-15 00:00:00') /
POW(10, 6)
   AND web100_log_entry.log_time < PARSE_UTC_USEC('2011-11-21 23:59:59') /
POW(10, 6)
   AND IS_EXPLICITLY_DEFINED(web100_log_entry.connection_spec.local_ip)
   AND IS_EXPLICITLY_DEFINED(web100_log_entry.snap.HCThruOctetsAcked)
   AND IS_EXPLICITLY_DEFINED(web100_log_entry.snap.SndLimTimeRwin)
   AND IS_EXPLICITLY_DEFINED(web100_log_entry.snap.SndLimTimeCwnd)
   AND IS_EXPLICITLY_DEFINED(web100_log_entry.snap.SndLimTimeSnd)
   AND IS_EXPLICITLY_DEFINED(project)
   AND project = 0
   AND IS_EXPLICITLY_DEFINED(connection_spec.data_direction)
   AND connection_spec.data_direction = 1
   AND IS_EXPLICITLY_DEFINED(web100_log_entry.is_last_entry)
   AND web100_log_entry.is_last_entry = True
   AND web100_log_entry.snap.HCThruOctetsAcked >= 8192
   AND (web100_log_entry.snap.SndLimTimeRwin +
        web100_log_entry.snap.SndLimTimeCwnd +
```

```
        web100_log_entry.snap.SndLimTimeSnd) >= 9000000
   AND (web100_log_entry.snap.SndLimTimeRwin +
        web100_log_entry.snap.SndLimTimeCwnd +
        web100_log_entry.snap.SndLimTimeSnd) < 3600000000
   AND IS_EXPLICITLY_DEFINED(web100_log_entry.snap.CongSignals)
   AND web100_log_entry.snap.CongSignals > 0
   AND (web100_log_entry.snap.State == 1
        OR (web100_log_entry.snap.State >= 5
            AND web100_log_entry.snap.State <= 11))
   GROUP BY ips

  );
```

3. % of packets retransmitted

The following query extracts the median % of packets retransmitted values of the country(here United States) over a period of one week(here Nov 15 – Nov 21)

The condition to get packets retransmitted is :
(web100_log_entry.snap.SegsRetrans/web100_log_entry.snap.DataSegsOut)
We have further averaged it over IPs before calculating the median
for the country as a whole.

To convert to percentage I have multiplied the final value by 100

Final Query :

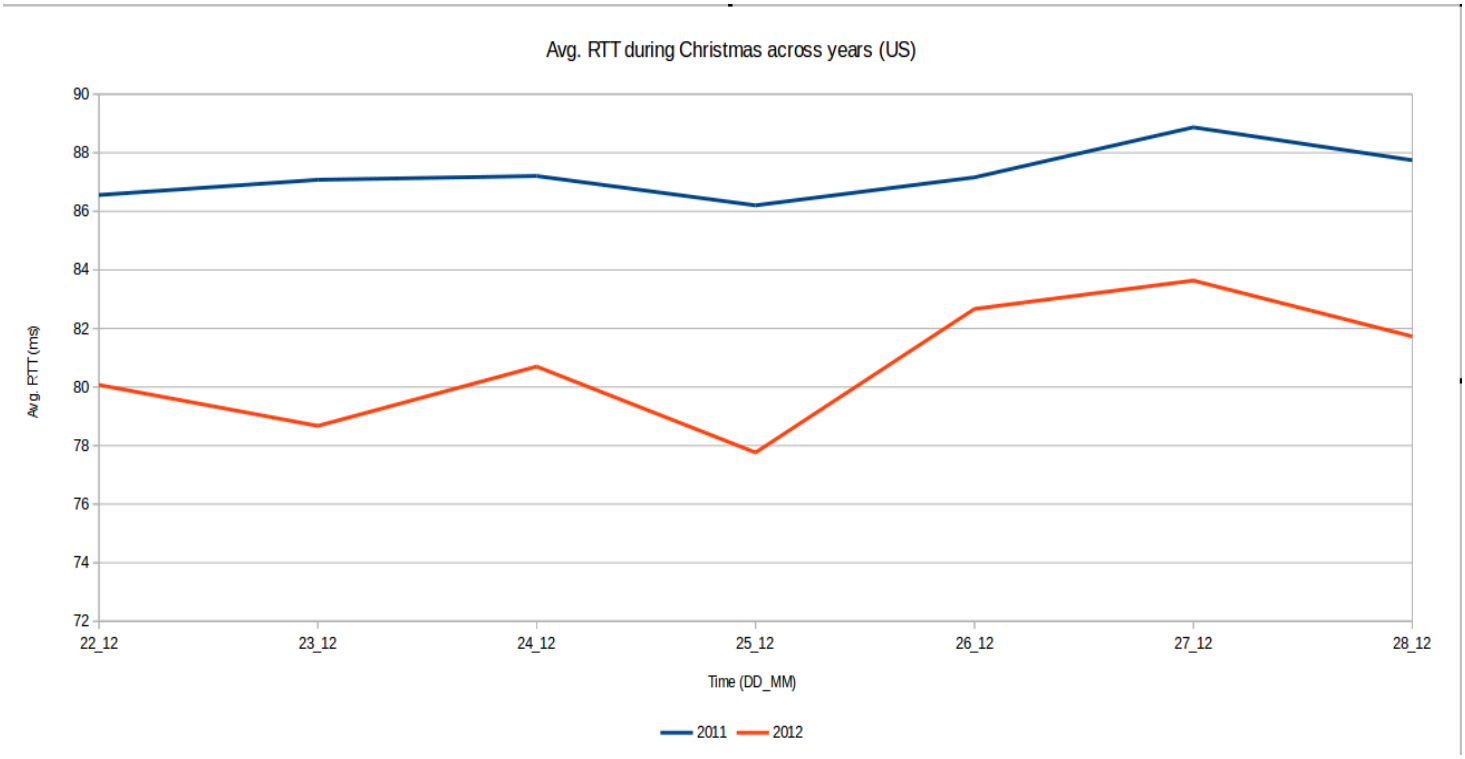
```
SELECT percentile_cont(0.5) OVER (ORDER BY retrans)
FROM (SELECT web100_log_entry.connection_spec.remote_ip AS ips,
AVG(web100_log_entry.snap.SegsRetrans/web100_log_entry.snap.DataSegsOut) AS
retrans
FROM [measurement-lab:m_lab.2011_11] WHERE
   IS_EXPLICITLY_DEFINED(web100_log_entry.connection_spec.remote_ip)
   AND IS_EXPLICITLY_DEFINED(connection_spec.client_geolocation.country_name)
   AND connection_spec.client_geolocation.country_name='India'
   AND IS_EXPLICITLY_DEFINED(web100_log_entry.log_time)
   AND web100_log_entry.log_time > PARSE_UTC_USEC('2011-11-15 00:00:00') /
POW(10, 6)
   AND web100_log_entry.log_time < PARSE_UTC_USEC('2011-11-21 23:59:59') /
POW(10, 6)
   AND IS_EXPLICITLY_DEFINED(web100_log_entry.connection_spec.local_ip)
   AND IS_EXPLICITLY_DEFINED(web100_log_entry.snap.HCThruOctetsAcked)
   AND IS_EXPLICITLY_DEFINED(web100_log_entry.snap.SndLimTimeRwin)
   AND IS_EXPLICITLY_DEFINED(web100_log_entry.snap.SndLimTimeCwnd)
   AND IS_EXPLICITLY_DEFINED(web100_log_entry.snap.SndLimTimeSnd)
   AND IS_EXPLICITLY_DEFINED(project)
   AND project = 0
   AND IS_EXPLICITLY_DEFINED(connection_spec.data_direction)
   AND connection_spec.data_direction = 1
   AND IS_EXPLICITLY_DEFINED(web100_log_entry.is_last_entry)
   AND web100_log_entry.is_last_entry = True
   AND web100_log_entry.snap.HCThruOctetsAcked >= 8192
   AND (web100_log_entry.snap.SndLimTimeRwin +
        web100_log_entry.snap.SndLimTimeCwnd +
        web100_log_entry.snap.SndLimTimeSnd) >= 9000000
   AND (web100_log_entry.snap.SndLimTimeRwin +
        web100_log_entry.snap.SndLimTimeCwnd +
        web100_log_entry.snap.SndLimTimeSnd) < 3600000000
   AND IS_EXPLICITLY_DEFINED(web100_log_entry.snap.SegsRetrans)
   AND IS_EXPLICITLY_DEFINED(web100_log_entry.snap.DataSegsOut)
   AND web100_log_entry.snap.DataSegsOut > 0
   AND (web100_log_entry.snap.State == 1
        OR (web100_log_entry.snap.State >= 5
```

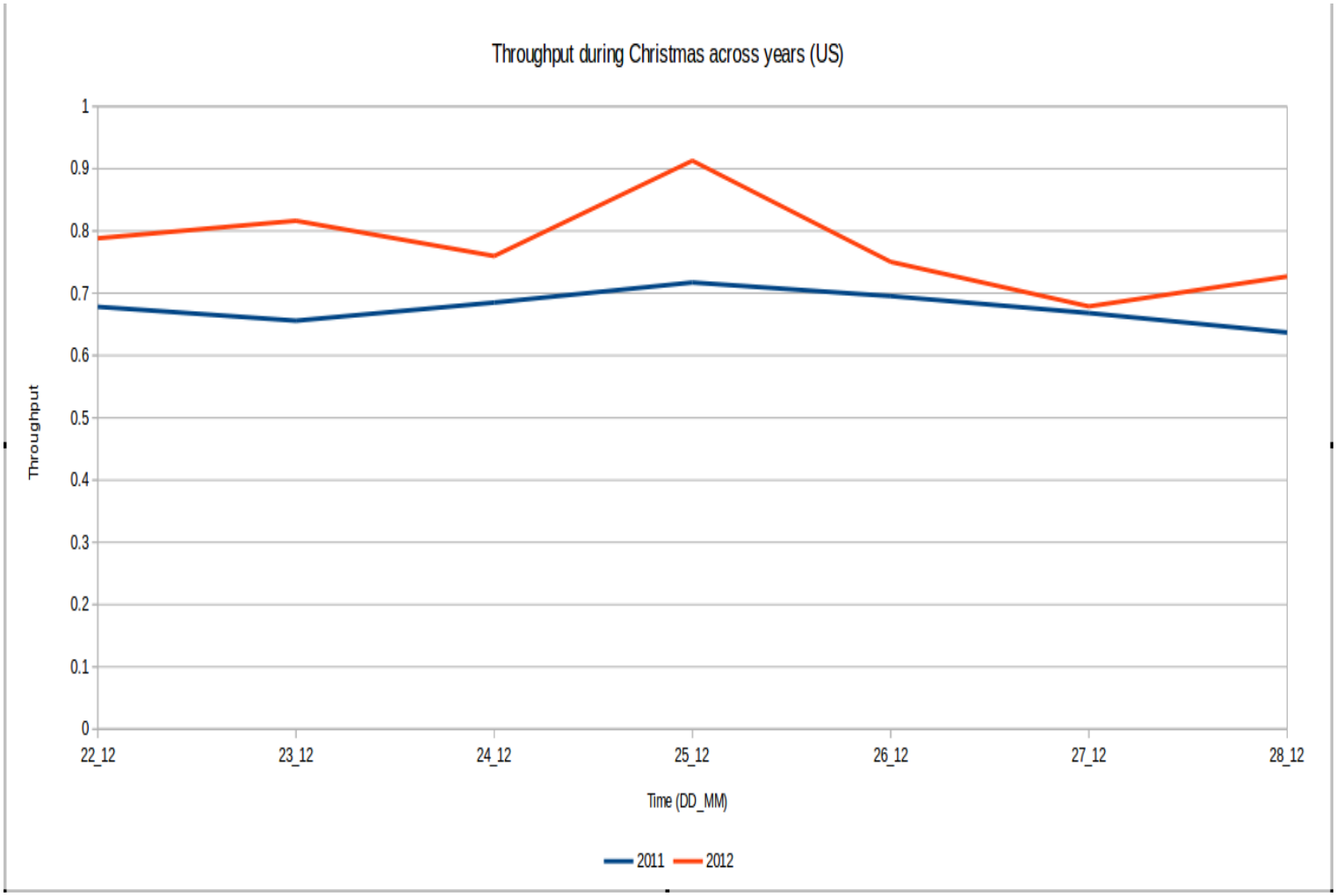
```
        AND web100_log_entry.snap.State <= 11))
GROUP BY ips
   );
```

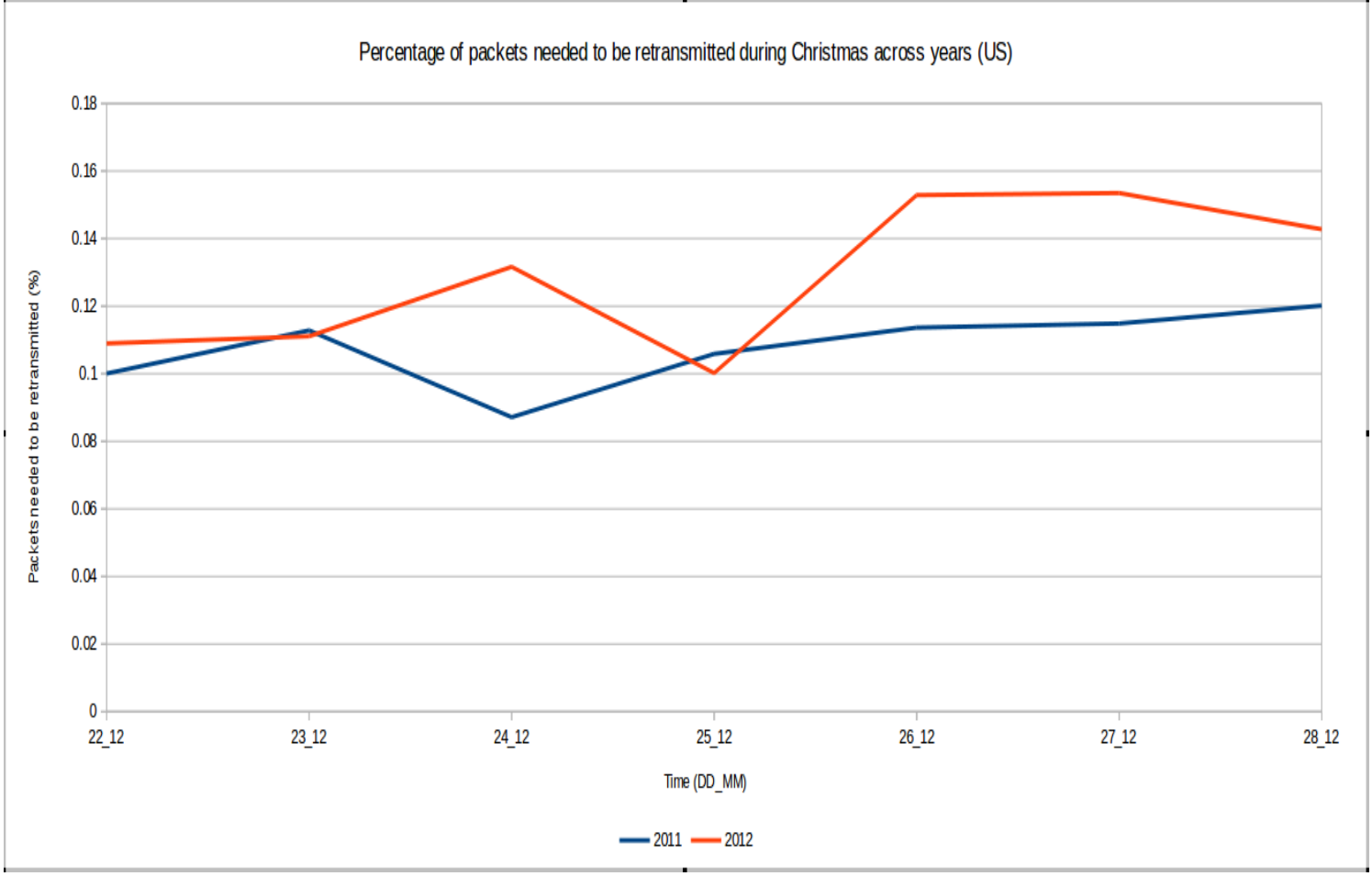
Results :

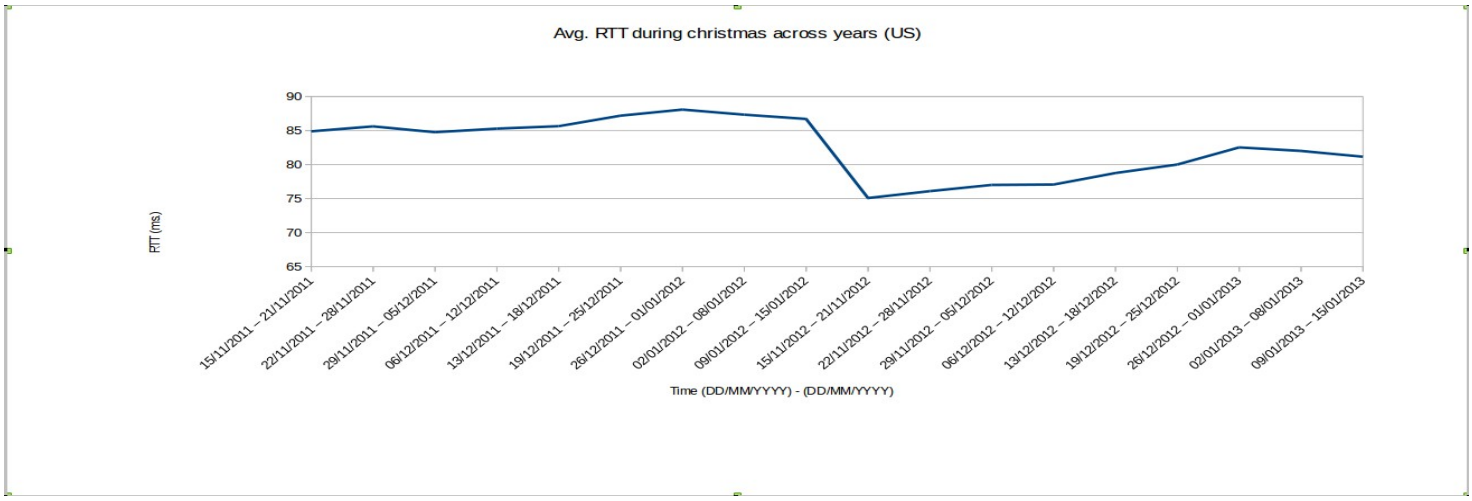
The graphs for these different parameters are attached below.

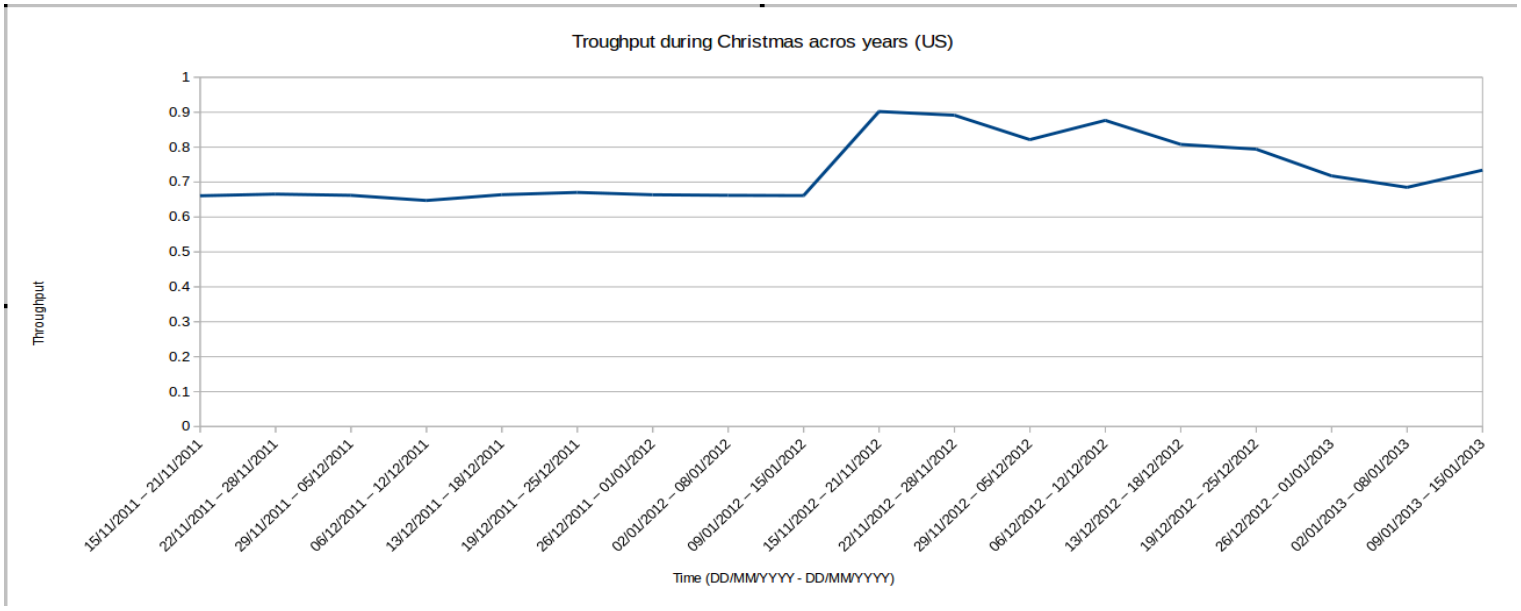
As we can see that there is a drop in Average RTT and a rise in Download Throughput only on the christmas day every year. Also while the % of packets that needed to be retransmitted drop in 2012, this is not the case in 2011. This fact warrants further inspection whether it is an anomaly or has some cause. Thus we can say that during Christmas people prefer to be with their family and friends in person rather than connect with them virtually.

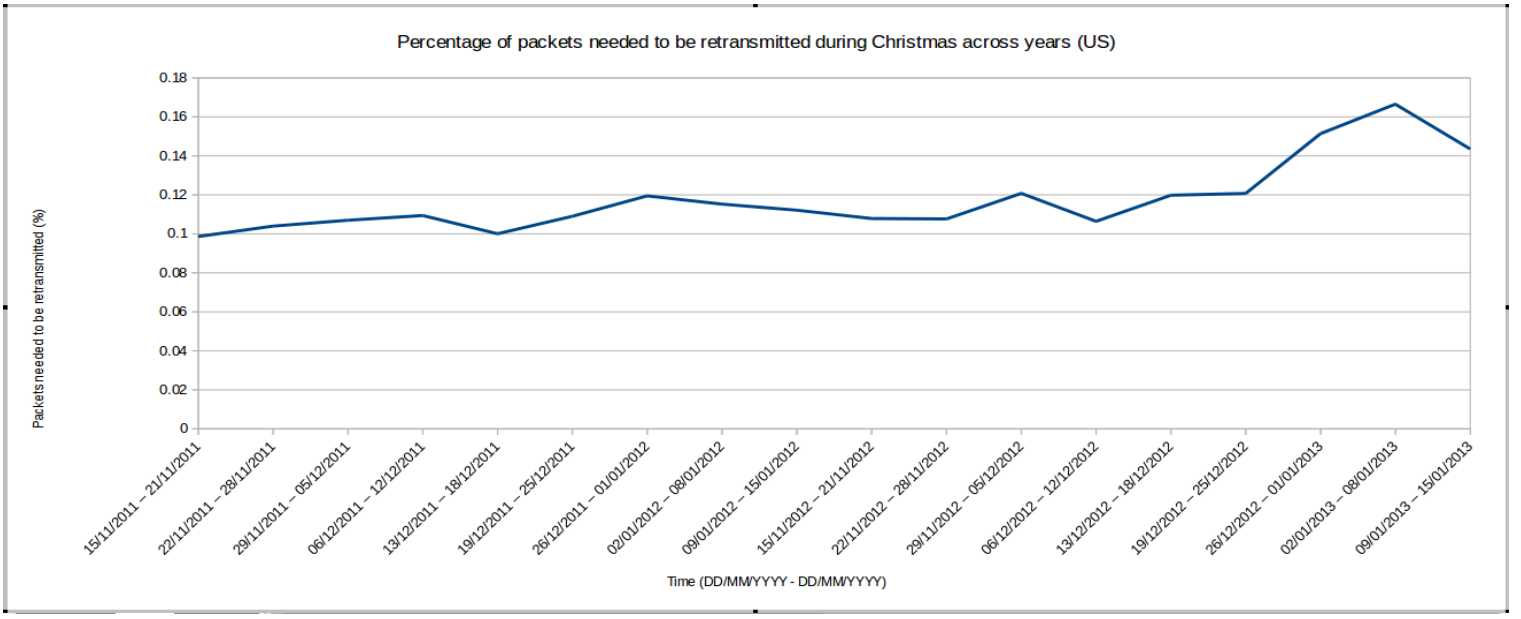



Throughput during Christmas across years (US)
Throughput
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
22_12
23_12
24_12
25_12
26_12
27_12
28_12
Time (DD_MM)
2011
2012


Percentage of packets needed to be retransmitted during Christmas across years (US)
Packets needed to be retransmitted (%)
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
22_12
23_12
24_12
25_12
26_12
27_12
28_12
Time (DD_MM)
2011
2012


Avg. RTT during christmas across years (US)
RTT (ms)
90
85
80
75
70
65
15/11/2011 – 21/11/2011
22/11/2011 – 28/11/2011
29/11/2011 – 05/12/2011
06/12/2011 – 12/12/2011
13/12/2011 – 18/12/2011
19/12/2011 – 25/12/2011
26/12/2011 – 01/01/2012
02/01/2012 – 08/01/2012
09/01/2012 – 15/01/2012
15/11/2012 – 21/11/2012
22/11/2012 – 28/11/2012
29/11/2012 – 05/12/2012
06/12/2012 – 12/12/2012
13/12/2012 – 18/12/2012
19/12/2012 – 25/12/2012
26/12/2012 – 01/01/2013
02/01/2013 – 08/01/2013
09/01/2013 – 15/01/2013
Time (DD/MM/YYYY) - (DD/MM/YYYY)


Troughput during Christmas acros years (US)
Throughput
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
15/11/2011 – 21/11/2011
22/11/2011 – 28/11/2011
29/11/2011 – 05/12/2011
06/12/2011 – 12/12/2011
13/12/2011 – 18/12/2011
19/12/2011 – 25/12/2011
26/12/2011 – 01/01/2012
02/01/2012 – 08/01/2012
09/01/2012 – 15/01/2012
15/11/2012 – 21/11/2012
22/11/2012 – 28/11/2012
29/11/2012 – 05/12/2012
06/12/2012 – 12/12/2012
13/12/2012 – 18/12/2012
19/12/2012 – 25/12/2012
26/12/2012 – 01/01/2013
02/01/2013 – 08/01/2013
09/01/2013 – 15/01/2013
Time (DD/MM/YYYY - DD/MM/YYYY)


Percentage of packets needed to be retransmitted during Christmas across years (US)
Packets needed to be retransmitted (%)
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
15/11/2011 – 21/11/2011
22/11/2011 – 28/11/2011
29/11/2011 – 05/12/2011
06/12/2011 – 12/12/2011
13/12/2011 – 18/12/2011
19/12/2011 – 25/12/2011
26/12/2011 – 01/01/2012
02/01/2012 – 08/01/2012
09/01/2012 – 15/01/2012
15/11/2012 – 21/11/2012
22/11/2012 – 28/11/2012
29/11/2012 – 05/12/2012
06/12/2012 – 12/12/2012
13/12/2012 – 18/12/2012
19/12/2012 – 25/12/2012
26/12/2012 – 01/01/2013
02/01/2013 – 08/01/2013
09/01/2013 – 15/01/2013
Time (DD/MM/YYYY - DD/MM/YYYY)