Abstract

A study was performed at the Stevens Institute of Technology which compared internet usage among various categories. The goal of the study was threefold: to determine whether wired or wireless internet was used more by students, to determine whether certain days of the week had greater traffic, specifically whether there was greater traffic on weekends, and to determine whether there was greater traffic at certain times of the day, specifically whether there was greater traffic in the afternoon and night compared to the morning hours. The study revealed interesting results. It was shown that for all dorms, wireless internet was used less than wired internet. No significant difference was detected in internet usage among days of the week. Finally, it was shown that there was significantly less internet usage in the morning than in the afternoon or nighttime.

Introduction

The purpose of this statistical study was to analyze the internet usage and traffic patterns in the student housing at Stevens Institute of Technology. To do so, the internet traffic from all of the dorms would have to be recorded and analyzed. To accomplish the goals of our study, we would need to be able to get an hour-by-hour picture of the internet traffic every day. By knowing how much data was being transmitted during ever hour of every day, we were able to compare different times of the day to each other and also compare the internet traffic of different days to each other. This allowed us to perform statistical tests to see if there were significant differences in the traffic between different times of day and between different days of the week. We expected there to be greater traffic on weekends, when students would have less work and therefore more free time to use the internet, and greater traffic at night when compared to the morning, because many students are either sleeping or in class during the morning hours.

Methods and Data

The process of which we gathered the data was all electronically. The data was sent every night to Giovanni Gaccione as a XML file. This file contained 14 files each with 289 lines of data. This data was separated into three parts. Part one was time, then the bytes/sec in, last was bytes/sec out. The data was broken down into 5 minute intervals. Giovanni took that file ran it through a parsing program that he wrote. The program took the files in XML and turned it

into a csv file. This happened to all 14 files every night at midnight. At the end we had close to 500 files. Another program ran to combine all the files and grouped by building. We then ran through the data or problems and inconsistencies. An example of this for a few hours of Oct. 19th the system was giving non numeric data. The group then decided that five minute intervals were too small. We then ran a Macro that grouped the data points into one hour sections. After this the data was sent out to the other members for analysis. After the data was organized into Excel spreadsheets, it was inputted into R statistical computing software and analyzed. The primary method of comparison was to use t-tests to test for significant differences in the mean internet traffic among the categories we defined. The first category was to divide each dorm's usage of wired and wireless internet. The second category was to split the total usage (wired and wireless combined) into the different days of the week. The third and final category was to divide the total usage (wired and wireless combined) into morning, afternoon, evening and late-night. Morning was defined to be 5am to 10:59am. Afternoon was defined to be 11am to 4:49pm. Evening was defined to be 5pm to 10:59pm. Late-night was defined to be 11pm to 4:49am.

Results

The results of all the tests were calculated and were interesting when compared to the hypotheses. Appendix Two shows the t-test results for the first group of tests. This group of tests compared wireless internet usage to wired internet usage for all seven of the dorms. For all seven dorms, the wireless usage was shown to be significantly lower than the wired usage. In addition, the differences in means were very large, the wired usage was anywhere from two to ten times greater than the wireless usage, so not only is the difference statistically significant, but it is also a meaningful difference. Appendix Three shows the t-test results for the second group of tests. This group of tests compared the mean usage for all of the dorms on each day of the week compared to the overall mean usage. What was interesting was that the tests showed that no days had a mean usage significantly different than the mean. Thus, our hypothesis that the usage would be greater on weekends was rejected. Appendix Four shows the t-test results for the third group of tests. This group of tests compared internet usage at various times of the day with the mean usage. In this instance, the tests seemed to confirm, or at least not reject, our hypothesis that there would be greater traffic at night than in the morning. The morning time period was shown to have traffic significantly less than the mean. This difference was also meaningful, as the means for the morning period were often half as much as the overall mean usage. The other three periods were shown to have traffic that was not significantly different than the mean.

Appendix One

Building A, wired:

mean traffic:	1338373
Monday traffic:	1604197
Tuesday traffic:	1757348
Wednesday traffic:	1229878
Thursday traffic:	1090288
Friday traffic:	1231845
Saturday traffic:	1259045
Sunday traffic:	1204179
Morning traffic:	0635879
Afternoon traffic:	1567474
Evening traffic:	1658074
Late traffic:	1500632

Building A, wireless

mean traffic:	0385180
Monday traffic:	0526174
Tuesday traffic:	0356573
Wednesday traffic:	0469752
Thursday traffic:	0354636
Friday traffic:	0319983
Saturday traffic:	0267067
Sunday traffic:	0390195
Morning traffic:	0151293
Afternoon traffic:	0443334
Evening traffic:	0510348
Late traffic:	0431462

Building B, wired:

mean traffic:	0933986
Monday traffic:	0970167
Tuesday traffic:	1116224
Wednesday traffic:	1018417
Thursday traffic:	0762847
Friday traffic:	0779752
Saturday traffic:	0877948
Sunday traffic:	1016388
Morning traffic:	0554198
Afternoon traffic:	0923850
Evening traffic:	1250360
Late traffic:	1004698

Building B, wireless
mean traffic: 0485406
Monday traffic: 0519860
Tuesday traffic: 0566325
Wednesday traffic: 0556439
Thursday traffic: 0426492
Friday traffic: 0420015
Saturday traffic: 0356354
Sunday traffic: 0544235
Morning traffic: 0252045
Afternoon traffic: 0457090
Evening traffic: 0601969
Late traffic: 0632667

Building C, wired:
mean traffic: 2226004
Monday traffic: 2329368
Tuesday traffic: 2623959
Wednesday traffic: 2471978
Thursday traffic: 2063439
Friday traffic: 2029675
Saturday traffic: 1795949
Sunday traffic: 2228262
Morning traffic: 1355366
Afternoon traffic: 2186652
Evening traffic: 2685961
Late traffic: 2679904

Building C, wireless:
mean traffic: 0518840
Monday traffic: 0506922
Tuesday traffic: 0621179
Wednesday traffic: 0578023
Thursday traffic: 0459897
Friday traffic: 0540981
Saturday traffic: 0490616
Sunday traffic: 0437656
Morning traffic: 0273218
Afternoon traffic: 0589239
Evening traffic: 0595927
Late traffic: 0618526

Building D, wired:

mean traffic:	0940631
Monday traffic:	1104711
Tuesday traffic:	1281803
Wednesday traffic:	0872661
Thursday traffic:	0974713
Friday traffic;	0940907
Saturday traffic:	0848395
Sunday traffic:	0895936
Morning traffic:	0433394
Afternoon traffic:	1030682
Evening traffic:	1226443
Late traffic:	1070862

Building D, wireless:

mean traffic:	0232013
Monday traffic:	0262036
Tuesday traffic:	0211215
Wednesday traffic:	0240950
Thursday traffic:	0255734
Friday traffic:	0268880
Saturday traffic:	0177843
Sunday traffic:	0196666
Morning traffic:	0106967
Afternoon traffic:	0261524
Evening traffic:	0275969
Late traffic:	0284248

Building E, wired:

mean traffic:	2736774
Monday traffic:	3246202
Tuesday traffic:	3308035
Wednesday traffic:	2461840
Thursday traffic:	3043339
Friday traffic:	2458156
Saturday traffic:	1988153
Sunday traffic:	2538071
Morning traffic:	1749143
Afternoon traffic:	2883202
Evening traffic:	2883202
Late traffic:	3029265

Building E, wireless

mean traffic:	0220350
Monday traffic:	0290500
Tuesday traffic:	0251203
Wednesday traffic:	0246150
Thursday traffic:	0218306
Friday traffic:	0245362
Saturday traffic:	0180187
Sunday traffic:	0108421
Morning traffic:	0095698
Afternoon traffic:	0238270
Evening traffic:	0266755
Late traffic:	0279323

Building F, wired

mean traffic:	1454148
Monday traffic:	1539098
Tuesday traffic:	1669985
Wednesday traffic:	1441137
Thursday traffic:	1541839
Friday traffic:	1205513
Saturday traffic:	1508460
Sunday traffic:	1230649
Morning traffic:	0753407
Afternoon traffic:	1453307
Evening traffic:	1777971
Late traffic:	1824397

Building F, wireless

mean traffic:	0594082
Monday traffic:	0643501
Tuesday traffic:	0931012
Wednesday traffic:	0651830
Thursday traffic:	0614172
Friday traffic:	0463007
Saturday traffic:	0314976
Sunday traffic:	0523886
Morning traffic:	0294939
Afternoon traffic:	0585958
Evening traffic:	0698761
Late traffic:	0798366

Building G, wired

mean traffic:	2434864
Monday traffic:	2999389
Tuesday traffic:	2875886
Wednesday traffic:	2766354
Thursday traffic:	2395189
Friday traffic:	2139794
Saturday traffic:	1623838
Sunday traffic:	2222221
Morning traffic:	1353150
Afternoon traffic:	2705124
Evening traffic:	2766664
Late traffic:	2908279

Building G, wireless

mean traffic:	0264697
Monday traffic:	0291261
Tuesday traffic:	0307411
Wednesday traffic:	0304691
Thursday traffic:	0224163
Friday traffic:	0213891
Saturday traffic:	0281405
Sunday traffic:	0228513
Morning traffic:	0131596
Afternoon traffic:	0299117
Evening traffic:	0307527
Late traffic:	0322836

Appendix Two

data: AWtraffic
t = -72.2009, df = 1006, p-value < 2.2e-16
alternative hypothesis: true mean is less than 1338373
99 percent confidence interval:
     -Inf 415941.7
sample estimates:
mean of x
 385180.3

data: BWtraffic
t = -30.0124, df = 1006, p-value < 2.2e-16
alternative hypothesis: true mean is less than 933986
99 percent confidence interval:
     -Inf 520232.2
sample estimates:
mean of x
   485406

data: CWtraffic
t = -113.084, df = 1006, p-value < 2.2e-16
alternative hypothesis: true mean is less than 2226004
99 percent confidence interval:
     -Inf 554016.5
sample estimates:
mean of x
 518840.9

data: DWtraffic
t = -97.0343, df = 1006, p-value < 2.2e-16
alternative hypothesis: true mean is less than 940631
99 percent confidence interval:
     -Inf 249029.2
sample estimates:
mean of x
 232013.4

data: EWtraffic
t = -308.2207, df = 1006, p-value < 2.2e-16
alternative hypothesis: true mean is less than 2736774
99 percent confidence interval:
     -Inf 239374.3
sample estimates:
mean of x
 220350.8

```
data:  FWtraffic
t = -47.3984, df = 1006, p-value < 2.2e-16
alternative hypothesis: true mean is less than 1454148
99 percent confidence interval:
     -Inf 636362.8
sample estimates:
mean of x
 594082.8

data:  GWtraffic
t = -252.478, df = 1006, p-value < 2.2e-16
alternative hypothesis: true mean is less than 2434864
99 percent confidence interval:
     -Inf 284725.2
sample estimates:
mean of x
 264697.3
```

Appendix Three

data: monday
t = 0.5468, df = 13, p-value = 0.5937
alternative hypothesis: true mean is not equal to 1054668
95 percent confidence interval:
  618820.2 1785949.2
sample estimates:
mean of x
  1202385

data: tuesday
t = 0.8035, df = 13, p-value = 0.4361
alternative hypothesis: true mean is not equal to 1054668
95 percent confidence interval:
  679178.3 1874844.3
sample estimates:
mean of x
  1277011

data: wednesday
t = 0.1662, df = 13, p-value = 0.8705
alternative hypothesis: true mean is not equal to 1054668
95 percent confidence interval:
  587935.4 1599221.8
sample estimates:
mean of x
  1093579

data: thursday
t = -0.101, df = 13, p-value = 0.9211
alternative hypothesis: true mean is not equal to 1054668
95 percent confidence interval:
  510191.5 1550530.5
sample estimates:
mean of x
  1030361

data: friday
t = -0.5265, df = 13, p-value = 0.6074
alternative hypothesis: true mean is not equal to 1054668
95 percent confidence interval:
  505106.9 1388859.0
sample estimates:
mean of x
   946983

data: saturday
t = -1.1369, df = 13, p-value = 0.2761
alternative hypothesis: true mean is not equal to 1054668
95 percent confidence interval:
 475623.3 1234410.4
sample estimates:
mean of x
 855016.9

data: sunday
t = -0.3281, df = 13, p-value = 0.748
alternative hypothesis: true mean is not equal to 1054668
95 percent confidence interval:
 512852.8 1453615.5
sample estimates:
mean of x
 983234.1

Appendix Four

data: morning
t = -3.2979, df = 13, p-value = 0.002885
alternative hypothesis: true mean is less than 1054668
95 percent confidence interval:
     -Inf 835560.7
sample estimates:
mean of x
 581449.5

data: afternoon
t = 0.2524, df = 13, p-value = 0.8047
alternative hypothesis: true mean is not equal to 1054668
95 percent confidence interval:
  590542.3 1641575.2
sample estimates:
mean of x
  1116059

data: evening
t = 0.5646, df = 11, p-value = 0.5837
alternative hypothesis: true mean is not equal to 1054668
95 percent confidence interval:
  644113.5 1748548.8
sample estimates:
mean of x
  1196331

data: late
t = 0.7048, df = 13, p-value = 0.4934
alternative hypothesis: true mean is not equal to 1054668
95 percent confidence interval:
  668162.6 1815475.2
sample estimates:
mean of x
  1241819