Alex Wang
axw582

Network Connectivity Speeds on Top Websites

Introduction / Motivation:

The objective of this report is to better understand network connectivity and the factors that may affect the "speed" of a network connection. Specifically, there will be a focus on trends between network speeds and the top-level domain of selected websites. Top-level domains (i.e. .com, .edu, .amazon, .navy, .jp) are the "suffixes" to website domain names, and they can sometimes convey information about the purpose or nature of the domain/website. For example, a website "example.gov" will likely be affiliated in some way with the US government. Another website "example.edu" will likely be affiliated with a higher-level learning institution. Top-level domains can even be associated with specific countries or companies (i.e. "example.jp" or "example.amazon"). The source for the websites analyzed in this report can be found here: http://s3.amazonaws.com/alexa-static/top-1m.csv.zip.

Procedure:

(a) Gathering Data:

The data-gathering process will involve the use of the ping utility on the top 5000 most popular websites ordered by Amazon's Alexa and listed in the link above. Although the source above includes substantially more than 5000 websites, only the top 5000 will be analyzed due to time limitations. To obtain the data, I will type "wget http://s3.amazonaws.com/alexa-static/top-1m.csv.zip" which will store a .csv file "top1-m.csv" on my local machine. After typing "make" and using the command "./proj5 -n 5000 -f top-1m.csv", the top 5000 websites stored in "top1-m.csv" will be pinged, and the results will be stored in a file "output.txt". This output includes the website name (including the top-level domain name) and the min/avg/max/mdev of the rrt (round-trip time) as reported by "ping [website-name] -q". In order to account for potential outlier rtts, each website will be pinged 5 times in total.

(b) Analyzing Data:

Analysis of the data will involve a program "sort.cpp", which will group websites with the same top-level domain name together. These groups will then have their associated data (gathered in part (a)) averaged out. A weakness to this approach is that some top-level domains only appear once or twice in the top 5000 websites, and therefore, their associated measurements are less reliable than those groups with a larger sample size.

To analyze the data, I will use the following commands:

- ./sort -f output.txt -m 0 >> min
- ./sort -f output.txt -m 1 >> avg
- ./sort -f output.txt -m 2 >> max
- ./sort -f output.txt -m 3 >> mdev

The file “min” will store a chart showing the fastest rtt (round-trip time) captured in milliseconds (ms) and a bar representing the relative speed of each TLD (top-level domain) grouping. The files “avg”, “max”, and “mdev” will show the average TLD, maximum TLD, and the mean deviation of the TLD, respectively. At this point, these four files (“min”, “avg”, “max”, and “mdev”) will be ordered in alphabetical order by TLD. Using an awk/sed command, these files will be reordered into increasing rtt order. This means the first line will store the fastest rtt captured, and the last line will store the slowest rtt captured. The awk command mentioned above is shown below, and it is called 4 times for the min/avg/max/mdev files respectively:

- awk '{printf "%.6f %s\n", $3, $0}' min | sort -n -k1,1 | sed -E -e 's/^[0-9\.]+ //' | sed -E -e 's/.*(-nan)//' | sed '/^$/d' >> orderedmin.txt
- awk '{printf "%.6f %s\n", $3, $0}' avg | sort -n -k1,1 | sed -E -e 's/^[0-9\.]+ //' | sed -E -e 's/.*(-nan)//' | sed '/^$/d' >> orderedavg.txt
- awk '{printf "%.6f %s\n", $3, $0}' max | sort -n -k1,1 | sed -E -e 's/^[0-9\.]+ //' | sed -E -e 's/.*(-nan)//' | sed '/^$/d' >> orderedmax.txt
- awk '{printf "%.6f %s\n", $3, $0}' mdev | sort -n -k1,1 | sed -E -e 's/^[0-9\.]+ //' | sed -E -e 's/.*(-nan)//' | sed '/^$/d' >> orderedmdev.txt
- (Since I have relatively little experience with awk/sed, I found some help here: https://unix.stackexchange.com/questions/63299/sort-a-file-based-on-length-of-the-column-row).

Since the awk command is somewhat complex, I will break it down into parts. The first portion “awk … [FILENAME]” creates a column containing the relative lengths of the rtt for each TLD. Next, “sort …” sorts the lines by that first column. Then, the remaining sed commands just clean up the information created to order the lines in this way. At this point all the information should be fully ordered. A screenshot of the “min” file before and after ordering can be found in Appendix (a) and (b).

In addition to all tools mentioned above, I also used basic vim functionality and Google Docs ctrl+f in order to discern any interesting patterns/trends in the data collected.

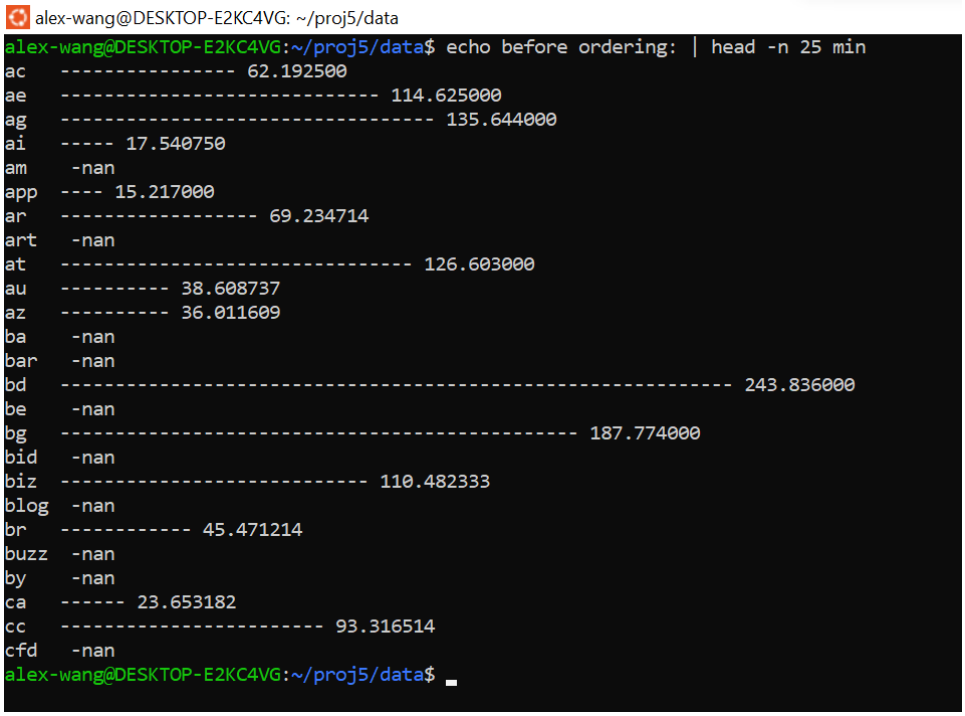
Results:

The following results are based on an analysis of pings from 5000 websites, each pinged 5 times. The following conclusions are based on no less than 5000 websites * 5 pings * 4 measurements = 100 000 data points:

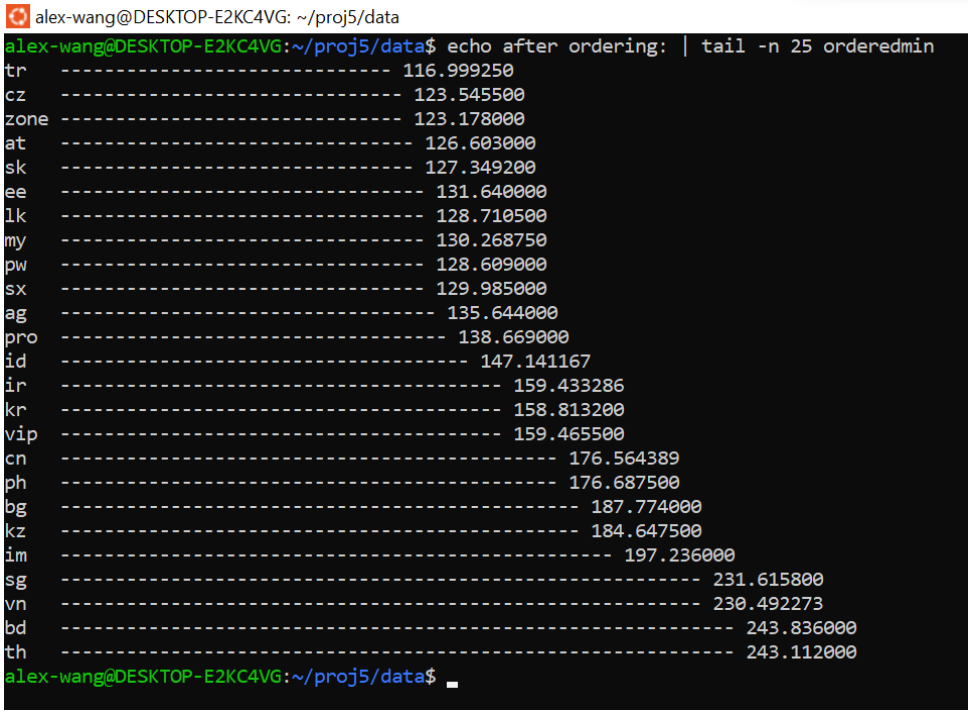
1. Based on the output stored in orderedmin (see Appendix (c)), the tdl ".ve" had the fastest average rtt (round-trip time) recorded by ping at 12.044 ms, followed by ".tube" and ".ro". These three TLDs are associated with Venezuela, a Latin American telecommunications company, and Romania, respectively. This is notable because many of the fastest websites found by proj5.cpp and sort.cpp are also some of the most reliably fast websites in the dataset as calculated in orederedmdev (see Appendix (f)).
2. Based on the last few rows of orderedavg.txt (see Appendix (d)) and orderedmdev.txt (see Appendix (e)), websites with the most inconsistent rtts were not necessarily the slowest. In fact, based on the last 20 lines of orderedavg.txt and orderedmdev.txt, only 6 of the websites with the slowest rtts were included in the websites with the most unreliable rtts.
3. Based on the output in orderedavg.txt (see Appendix (d)), TLDs associated with countries that are geographically distant from the U.S. tend to be "slower" than countries closer to the U.S. TLDs such as ".bd" (Bangladesh), ".th" (Thailand), and ".vn" (Vietnam) had the top 3 slowest rtts obtained by proj5.cpp. On the other hand, countries such as ".ve" (Venezuela) and ".pe" (Peru) were significantly faster.
4. Based on the results in orderedavg.txt out of the commonly-used TLDs (see Appendix (d) and (g) source: https://en.wikipedia.org/wiki/Generic_top-level_domain), the fastest historic generic TLDs are ".gov", ".edu", ".org", ".com", and ".net", listed in decreasing rtts. These results seem to indicate that TLDs associated with the government (".gov") and higher education (".edu") are generally faster than those for commercial purposes (".com" and ".net").
5. Based on the results from orderedmax.txt (see Appendix (e)), TLDs with the longest captured rtts tended to also have the longest minimum rtts and the longest average rtts (see Appendix (c) and (d)).
6. Based on the first listed results in orderedmdev.txt (see Appendix (f)), the TLDs with the most consistent rtts (those with the smallest "mdev" values) were not necessarily the fastest. In fact, only 3 TLDs were found in common between the first 20 lines of orderedavg.txt and orderedmdev.txt. This suggest very little overlap between the two groups.
7. Based on the data in orderedavg.txt, the vast majority (109 out of 113 TDLs) have a rtt less than 200.00 ms. Additionally, no TDLs measured had a rtt of greater than 0.300 seconds.

Appendix:

a)



b)



c) orderedmin.txt:

```
ve   ---- 12.044000
tube ---- 12.719000
```

```
ro   ---- 13.064500
gg   ---- 13.522500
hr   ---- 13.928000
so   ---- 13.979000
men  ---- 14.023000
wf   ---- 14.041000
ly   ---- 14.083500
run  ---- 14.094000
lat  ---- 14.445000
club ---- 14.514750
site ---- 14.545500
su   ---- 14.575000
cx   ---- 14.625000
life ---- 14.643000
app  ---- 15.217000
gov  ----- 16.229882
ai   ----- 17.540750
pe   ----- 17.739800
dev  ----- 19.594250
fm   ----- 19.861000
xyz  ----- 19.984143
nl   ------ 21.469600
live ------ 21.703500
re   ------ 23.081500
ca   ------ 23.653182
us   ------- 24.231667
edu  ------- 25.463160
mx   -------- 31.350000
wiki --------- 34.173333
co   --------- 34.392633
fi   --------- 34.730667
link --------- 34.765500
az   ---------- 36.011609
au   ---------- 38.608737
uk   ---------- 39.622346
org  ----------- 40.189852
to   ----------- 41.286000
xxx  ----------- 42.022250
tv   ----------- 42.984257
br   ------------ 45.471214
me   ------------ 45.629591
li   ------------ 46.470750
rs   ------------ 47.549750
fr   ------------ 47.822643
io   ------------ 47.851943
ua   ------------- 51.305167
info -------------- 54.146222
```

```
lol  -------------- 54.191667
gr   -------------- 54.363167
is   -------------- 54.818333
moe  --------------- 56.426000
es   --------------- 58.231222
pk   --------------- 59.560600
ac   ---------------- 62.192500
il   ---------------- 62.307667
no   ----------------- 65.842500
in   ----------------- 66.755870
it   ----------------- 67.071400
top  ------------------ 68.312167
ar   ------------------ 69.234714
com  ------------------ 69.470126
hu   ------------------ 70.584500
la   ------------------ 71.408667
de   ------------------- 74.966966
net  ------------------- 75.428899
one  -------------------- 76.739000
ch   -------------------- 77.453667
name -------------------- 77.825000
cl   -------------------- 78.492000
fun  -------------------- 79.085750
hk   ---------------------- 84.079333
pl   ---------------------- 88.749083
tw   ---------------------- 91.248833
jp   ---------------------- 91.899255
cc   ----------------------- 93.316514
qa   ------------------------ 99.600000
ru   ------------------------- 100.000817
news ------------------------- 100.782500
pt   ------------------------- 106.771000
sa   -------------------------- 108.484333
nz   -------------------------- 108.795000
desi -------------------------- 109.453000
biz  -------------------------- 110.482333
eg   -------------------------- 112.403000
dk   -------------------------- 113.284000
ae   -------------------------- 114.625000
tr   -------------------------- 116.999250
zone -------------------------- 123.178000
cz   -------------------------- 123.545500
at   -------------------------- 126.603000
sk   -------------------------- 127.349200
pw   -------------------------- 128.609000
lk   -------------------------- 128.710500
sx   -------------------------- 129.985000
```

```
my   ---------------------------------- 130.268750
ee   ---------------------------------- 131.640000
ag   ---------------------------------- 135.644000
pro  ---------------------------------- 138.669000
id   ------------------------------------ 147.141167
kr   ---------------------------------------- 158.813200
ir   ---------------------------------------- 159.433286
vip  ---------------------------------------- 159.465500
cn   ------------------------------------------------ 176.564389
ph   ------------------------------------------------ 176.687500
kz   ------------------------------------------------- 184.647500
bg   ------------------------------------------------- 187.774000
im   ---------------------------------------------------- 197.236000
vn   ------------------------------------------------------------ 230.492273
sg   ------------------------------------------------------------ 231.615800
th   ------------------------------------------------------------------ 243.112000
bd   ------------------------------------------------------------------ 243.836000
```

d) orderedavg.txt:

```
tube ---- 13.581000
gg   ---- 13.856000
run  ---- 14.469000
site ---- 15.079500
hr   ---- 15.108000
life ---- 15.346000
men  ---- 15.373000
club ---- 15.400500
cx   ---- 15.767000
su   ---- 15.831000
ro   ----- 16.095250
wf   ----- 16.302000
ve   ----- 16.682000
gov  ----- 17.337235
ai   ----- 18.295250
lat  ----- 18.535000
pe   ----- 18.625600
ly   ----- 18.645000
app  ----- 19.574200
fm   ------ 20.674000
so   ------ 20.799000
dev  ------ 20.853500
live ------ 22.652167
nl   ------ 23.246700
re   ------ 23.770500
us   ------- 25.287000
ca   ------- 26.461818
```

```
edu  ------- 27.770840
xyz  -------- 31.450571
wiki --------- 35.300000
link ---------- 36.447500
az   ---------- 37.366304
fi   ---------- 39.676333
mx   ----------- 40.009833
uk   ----------- 41.269808
org  ----------- 42.678733
au   ----------- 42.960842
xxx  ----------- 43.519250
to   ------------ 44.146452
co   ------------ 44.492267
li   ------------ 47.384000
tv   ------------- 48.394029
br   ------------- 48.546964
fr   ------------- 48.625286
rs   ------------- 50.021250
me   ------------- 51.012955
io   ------------- 51.943226
info -------------- 54.656667
is   -------------- 55.550667
gr   -------------- 55.804000
moe  --------------- 57.259000
lol  --------------- 58.966333
pk   ---------------- 60.477600
es   ---------------- 60.481444
ac   ---------------- 62.713500
il   ---------------- 62.985000
ua   ----------------- 66.781833
no   ----------------- 67.047000
it   ----------------- 67.984800
in   ------------------ 68.553391
top  ------------------ 70.495500
com  ------------------- 72.458992
de   -------------------- 76.292103
one  -------------------- 77.387667
ar   -------------------- 77.755714
ch   -------------------- 78.847000
name -------------------- 79.083000
la   -------------------- 79.416333
net  -------------------- 79.516654
fun  --------------------- 80.534500
hu   --------------------- 80.812500
cl   ---------------------- 86.014800
hk   ---------------------- 87.493000
tw   ----------------------- 93.425083
```

```
jp   ----------------------- 95.692655
cc   ------------------------ 96.794686
pl   ------------------------ 99.124583
qa   ------------------------- 100.386000
ru   ------------------------- 101.812305
news ------------------------- 102.255000
pt   -------------------------- 107.941000
sa   --------------------------- 109.556000
nz   --------------------------- 109.626000
eg   ---------------------------- 113.695000
desi ---------------------------- 114.423000
ae   ---------------------------- 115.116000
biz  ---------------------------- 115.683000
dk   ----------------------------- 116.515000
tr   ----------------------------- 118.728875
zone ------------------------------ 124.691000
pw   ------------------------------- 129.252000
lk   ------------------------------- 129.365500
sk   ------------------------------- 129.821200
cz   ------------------------------- 130.756400
sx   ------------------------------- 131.162000
ee   -------------------------------- 133.225000
my   -------------------------------- 133.935250
ag   --------------------------------- 136.175000
pro  --------------------------------- 139.673667
id   ---------------------------------- 151.584667
at   ------------------------------------ 158.409500
ir   ------------------------------------ 161.436381
kr   ------------------------------------- 164.575600
vip  ------------------------------------- 170.982500
ph   -------------------------------------- 177.198000
cn   -------------------------------------- 180.559459
bg   --------------------------------------- 188.856000
kz   --------------------------------------- 189.443000
im   ---------------------------------------- 198.408000
sg   ------------------------------------------------------------ 232.388600
vn   ------------------------------------------------------------ 235.968636
bd   ------------------------------------------------------------- 244.088000
th   -------------------------------------------------------------- 249.325500
```

e) orderedmax.txt:

```
gg   ---- 14.570750
run  ---- 14.889000
tube ---- 15.649000
cx   ----- 16.814000
site ----- 16.944500
```

```
life ----- 16.973000
club ----- 17.238750
su   ----- 17.643000
hr   ----- 17.768000
men  ----- 18.063000
gov  ----- 19.372588
ai   ----- 19.537500
pe   ----- 19.898600
wf   ------ 20.989000
fm   ------ 21.731000
dev  ------ 23.428750
live ------- 24.317667
re   ------- 24.704500
ro   ------- 25.197750
nl   ------- 25.999600
us   ------- 27.757000
so   -------- 30.348000
app  -------- 31.184000
ve   -------- 31.336000
lat  --------- 32.511000
edu  --------- 34.574400
ly   --------- 35.333000
ca   --------- 35.407909
wiki ---------- 36.304667
az   ---------- 39.423739
link ----------- 41.598500
uk   ------------ 45.324423
fi   ------------ 45.508000
xyz  ------------ 46.393714
org  ------------ 46.986163
xxx  ------------ 46.988500
au   ------------ 47.491474
li   ------------- 48.905750
fr   ------------- 50.086071
to   ------------- 50.088419
mx   ------------- 51.456083
br   -------------- 52.396000
info -------------- 55.340333
is   --------------- 56.657000
co   --------------- 56.723033
rs   --------------- 57.104500
moe  --------------- 57.717000
gr   --------------- 57.904000
me   --------------- 59.280636
io   --------------- 59.802472
tv   --------------- 59.828543
pk   ---------------- 62.595400
```

```
ac   ---------------- 62.993500
il   ---------------- 63.696667
es   ----------------- 66.831556
it   ------------------ 69.676000
no   ------------------ 70.075750
in   ------------------- 72.096304
top  ------------------- 72.848583
lol  ------------------- 73.484000
ua   -------------------- 77.994333
com  -------------------- 78.024912
one  -------------------- 78.102000
de   -------------------- 78.275034
name -------------------- 79.599000
ch   --------------------- 81.099667
fun  --------------------- 83.422250
net  ---------------------- 86.463995
ar   ----------------------- 95.955714
cl   ------------------------ 97.067200
tw   ------------------------ 98.626583
hk   ------------------------ 99.615667
qa   ------------------------- 101.138000
la   ------------------------- 101.567333
cc   ------------------------- 102.242057
jp   ------------------------- 102.403836
news ------------------------- 103.980500
ru   -------------------------- 106.226805
pt   --------------------------- 109.124000
nz   --------------------------- 110.199000
pl   --------------------------- 110.340167
sa   ---------------------------- 112.655333
ae   ---------------------------- 116.113000
eg   ---------------------------- 117.535000
hu   ---------------------------- 118.760000
tr   ----------------------------- 123.418750
biz  ----------------------------- 124.558333
dk   ----------------------------- 127.050000
zone ----------------------------- 128.717000
pw   ----------------------------- 129.725000
lk   ----------------------------- 130.226500
desi ----------------------------- 131.344000
sx   ----------------------------- 132.060000
sk   ----------------------------- 135.559400
ag   ----------------------------- 137.585000
cz   ----------------------------- 138.925500
ee   ----------------------------- 139.089000
pro  ----------------------------- 141.310000
my   ----------------------------- 146.489250
```

```
ir   ---------------------------------------- 164.717238
id   ---------------------------------------- 168.100833
ph   ---------------------------------------- 178.042250
kr   ---------------------------------------- 182.717200
vip  ---------------------------------------- 187.357000
cn   ---------------------------------------- 187.499643
bg   ---------------------------------------- 189.694000
kz   ---------------------------------------- 197.599000
im   ---------------------------------------- 199.513000
sg   ---------------------------------------- 233.265400
bd   ---------------------------------------- 244.474000
vn   ---------------------------------------- 247.394455
at   ---------------------------------------- 257.051000
th   ----------------------------------------
268.392000
```

f) orderedmdev.txt:

```
ac   - 0.288500
run  - 0.302000
gg   - 0.401000
one  - 0.484667
nz   - 0.490000
info - 0.490111
ph   - 0.508250
bd   - 0.530000
il   - 0.559333
re   - 0.561000
sg   - 0.630600
name - 0.644000
pw   - 0.650000
is   - 0.676667
fm   - 0.683000
sx   - 0.738000
pt   - 0.744000
wiki - 0.786667
pe   - 0.802200
ai   - 0.817500
ae   - 0.844000
life - 0.852000
ag   - 0.854000
fr   - 0.855143
cx   - 0.878000
lk   - 0.885000
moe  - 0.936000
pro  - 0.972333
it   - 0.984900
```

```
su   - 0.998000
li   - 1.003500
club - 1.034500
live - 1.047500
site - 1.091500
gov  - 1.195059
tube - 1.217000
news - 1.218000
de   - 1.287207
az   - 1.295522
pk   - 1.299400
us   - 1.305333
hr   - 1.363000
gr   - 1.401667
ch   - 1.420000
dev  - 1.429000
qa   - 1.529000
im   - 1.530000
men  - 1.563000
fun  - 1.597500
no   - 1.652000
nl   - 1.660700
sa   - 1.691000
top  - 1.719583
eg   - 1.929000
xxx  - 1.967500
ir   - 1.969333
in   - 1.994435
uk   - 2.173462
ru   - 2.373634
wf   - 2.413000
tr   - 2.463000
org  - 2.589837
tw   - 2.766833
link - 2.800000
br   - 2.883964
ee   - 2.933000
sk   - 3.119200
to   - 3.323419
es   - 3.335556
com  - 3.343769
cc   - 3.400857
edu  - 3.578960
rs   - 3.679500
au   - 3.696789
fi   - 3.767333
jp   -- 4.031691
```

```
cn   -- 4.252108
net  -- 4.350644
ca   -- 4.653455
ro   -- 4.676500
io   -- 4.745358
kz   -- 4.996000
me   -- 5.153955
biz  -- 5.921667
cz   -- 5.929500
app  -- 6.134400
hk   -- 6.214000
my   -- 6.422500
cl   -- 6.640400
tv   -- 6.653514
vn   -- 6.805182
lat  -- 7.005000
dk   -- 7.286000
ve   -- 7.346000
lol  -- 7.427000
so   -- 7.704000
id   --- 8.361000
desi --- 8.478000
pl   --- 8.507583
co   --- 8.625733
mx   --- 8.783167
ly   --- 8.785000
kr   --- 9.462600
ua   --- 9.793667
ar   --- 9.884714
xyz  --- 11.399429
zone --- 11.674000
th   --- 11.848500
la   --- 11.857667
vip  ---- 12.839000
hu   ----- 19.111500
bg   --------- 33.625000
at   ------------- 50.635500
```

g)

Historical generic TLDs	
Domain	**Intended use**
`com`	Mainly for commercial entities, but unrestricted
`org`	Originally for organizations not clearly falling within the other gTLDs, now unrestricted
`net`	Originally for network infrastructures, now unrestricted
`edu`	Educational use, but now primarily for third-level colleges and universities
`gov`	Governmental use, but now primarily for US governmental entities and agencies
`mil`	Military use, but now primarily for US military only