PingER Host Searcher

 Increasing the coverage of PingER in developing countries has been difficult so far because it is hard to find hosts which are geographically located within those countries and do not block pings. The usual method of searching for hosts on Google using the top level domain of such countries has proven to be tedious and time consuming task. The various steps involved in this are:

1. Search for possible hosts on Google using the tld
2. Ping each host manually to check if it blocks pings
3. Using Visual Traceroute or GeoIPtool to confirm whether the host is geographically located in the desired country.

The PingER host searcher is an attempt to solve this problem. It completely automates the above procedure by:

1. It automatically downloads the results for the required country using its tld from Google. The default number of search results to download is 1000 but it is ocnfigurable and can be specified in multiples of 100 upto a maximum of 1000.
2. Using regular expressions and pattern matching it searches for hostnames in the results.
3. After elimination of any duplicate hostnames in the list it starts pinging them individually. At this stage the user can configure the number of pings he wants to send out to each host and the time-out value of the whole ping command. The default value is 10 sec timeout for 10 pings.
4. After the results of the pings come in, the program filters out hosts which block pings and also those with multiple hostnames for the same IP address, keeping one copy for a single IP address. It also stores the min_rtt for all the hosts in the filtered list.
5. Finally it checks the hosts in the filtered list on GeoIptool, again using pattern matching to show the top level domain, country, city and lat/long for each host. This information has no guarantee of being absolutely correct but on numerous occasions it was observed that it does possess a high degree of accuracy. The results at this stage can be configured by command line to be filtered optionally by either a threshold min_rtt or the top level domain of the results or both.

The program is available on SVN. You can download it by checking it out of svn using the command.

svn co file:///afs/slac.stanford.edu/g/scs/net/netmon/repo/svn/pinger/trunk/bin/HostSearcher.pl

```
akbar@pinger $ perl HostSearcher.pl

    usage: HostSearcher.pl --tld top_level_domain [--max_hosts] [--max_pings] [-max_time] [--filter] [--
pingserver] [--file]

    Options:
        --tld                  The two letters of the top level domain of the country.
        --max_hosts            Maximum number of hosts to download from Google ( >= 100 and <=1000 ) Default
is 1000
        --max_pings            Maximum number of pings to send to a host. Default is 10
        --max_time             Time before the ping to an individual host times out and returns. Default is 10
sec
        --filter               Filter out the results. Possible values are 'tld' , 'rtt' or 'off'. Default is
'off'
        --pingserver           Use a pingserver to ping rather than using the current host (To be implemented)
        --file                 File to store the results. Default is a file named tld.txt in the current folder

    example: HostSearcher.pl --tld pk
```

```
akbar@pinger $ perl HostSearcher.pl --tld ci --max_pings 2 --max_time 2 --max_hosts 100
--------------------------------------------------------------------
Searching Google for hosts........Done
--------------------------------------------------------------------
Parsing the downloaded data for hosts.......Done


====================================================================
The initial parsed list consists of 98 nodenames
A total of 96 unique nodenames were found
====================================================================


====================================================================
Pinging candidate nodes
====================================================================

Pinging host primature.ci.....Failed
Pinging host tchatche.ci.....Success!! IP-Address=81.199.127.114 Min-RTT=285.806 Country=(CI) City=Abidjan Lat
/Long= -4.0281, 5.3411
Pinging host vin.ci.....Success!! IP-Address=64.187.108.201 Min-RTT=68.542 Country=(US) City=Miami Lat/Long=
-80.1911, 25.7631
Pinging host www.ado.ci.....Success!! IP-Address=216.162.72.102 Min-RTT=91.041 Country=(CA) City=Canada Lat
/Long= -73.5833, 45.5
Pinging host www.aeria.ci.....Success!! IP-Address=213.136.96.35 Min-RTT=237.664 Country=(CI) City=Abidjan Lat
/Long= -4.0281, 5.3411
Pinging host www.ai3l.ci.....Success!! IP-Address=213.136.96.8 Min-RTT=244.688 Country=(CI) City=Abidjan Lat
/Long= -4.0281, 5.3411
Pinging host www.aip.ci.....Success!! IP-Address=213.186.33.2 Min-RTT=147.856 Country=France(FR) City=Roubaix
Lat/Long= 3.1667, 50.7
  |
  |
  |
  |
Pinging host www.uabobo.ci.....Success!! IP-Address=196.201.66.45 Min-RTT=537.522 Country=(CI) City= Lat/Long=
-5, 8
Pinging host www.ucacie.ci.....Success!! IP-Address=213.136.96.34 Min-RTT=237.347 !!Duplicate DNS hostname, IP
address already present in list!!
Pinging host www.ucocody.ci.....Success!! IP-Address=212.37.221.34 Min-RTT=148.288 Country=France(FR)
City=Paris Lat/Long= 2.3333, 48.8667
Pinging host www.uica.ci.....Success!! IP-Address=213.136.121.170 Min-RTT=280.823 Country=(CI) City= Lat/Long=
-5, 8
Pinging host www.uvci.ci.....Success!! IP-Address=213.136.96.12 Min-RTT=237.011 !!Duplicate DNS hostname, IP
address already present in list!!
====================================================================
A total of 42 nodes with unique ip addresses and satisying filter conditions(if any) were found
====================================================================
```

Criteria for Selecting Hosts

We manually choose sites based on the following criteria:

  * Does the min-RTT make sense compared to sites nearbye.
  * They are really in the country and not a proxy elsewhere (we use GeoIPTools to identify the country).
  * GeoIPTools has a city for the site.
  * The host is a web server - this often enables us to find out more about the site via the web.
  * The site is an educational or government site (in that order of preference).
  * Sites within a country are chosen for diversification, i.e. different cities, different uses (Education, Government, Commercial...), different IP
    network addresses.
  * We choose sites that appear to have better connectivity (e.g.lower RTT).
  * Where possible >= 2 sites/country (see Hosts per Country per Region).

Examples

```
HostSearcher.pl --tld lr --webonly --filter tld | tee pinger/hostsearcher/lr
```

```
grep '(LR' pinger/hostsearcher/lr | grep -v 'City= '
```

- Republic of Congo: Illustrates the difficulty of finding www hosts in some countries.
- More examples can be found at http://www-iepm.slac.stanford.edu/pinger/hostsearcher.

Usage

TLD	Date	Country	# Hosts searched	Success	#remaining hosts	Hits	Added	Comment
DJ	10/23/2010	Djibouti	272	No	0	1	0	with webonly, host already included, now pings
CG	10/23/2010	Congo Brazzaville	54	No	0	0	0	even without webonly
CZ	10/23/2010	Czech Republic	550	Yes	0	384	1	Hard to identify universities
GW	11/17/2009	Guinea Bissau	3	No	0	0	0	None matched filter
IR	10/23/2010	Iran	10	Yes	0	2	2	
MM	10/23/2010	Myanmar	111	No	0	13	0	Several .gov.my nodes all blocked, several nodes from domain 203.81.81
TJ	10/16/2010	Tajikistan	237	Yes	0	8	2	
CD	8/15/2009	DRC	30	No	4	0	0	
TD	11/17/2009	Chad	3	No	1	0	0	
RE	10/20/2010	Reunion	445	No	1	0	0	
MV	10/23/2010	Maldives	277	Yes	1	3	1	
IQ	10/24/2010	Iraq	19	No	0	0	0	Also tried search on Iraq universities including uobasrah.edu.iq, www.iraquniversity.net, uobaghdad.edu.iq
ZM	10/30/2010	Zambia	342	Yes	6	13	3	New nodes are outside Lusaka and have satellite
TLD	Date	Country	# Hosts	Success	# Remaining hosts	Hits	Added	Comments