Network Path Diagnostics **Nov 15 2009**

Deepak Konidena
Mandeep Mehta

Network path diagnostics builds up statistics on the traceroute utility. It calculates prominent statistics such as number of hops traversed, time taken, successful and unsuccessful journeys of a packet between source and destination, etc. This project tried to give a bigger and realistic picture of how the internet is connected.

Our program, npd.c when run takes four arguments in the form of

a) IP Address
b) Hostname
c) File containing IP Addresses
d) File containing Hostnames.

The function executeTraceroute(char *) executes the traceroute with the argument provided and writes the output to a file. We then use populateVal() to parse the output present in the file and extract values. We use two functions getDotCount() and isOnlyNumerical() to get the dot count of an individual string and whether the string contains only numerals to help us distinguish between ip addresses and hostnames present in the output.

We used strtok wherever possible to break the string into tokens. We also check for any invalid characters present in the input passed to traceroute. The function getDomain() is used to retrieve the domain name for a given hostname.

This project uses the longest prefix match for calculating the BGP prefixes instead of nexthop values. The BGP prefixes of the last non starred hop in the traceroute output determine how close the packet reached to the destination.

This project involved a substantial amount of file parsing. We faced problems in extracting the individual lines from the file in the form of linefeeds (\n). We had to prune the newline character at the end of each line. We made use of gethostbyname() and gethostbyaddr() functions to retrieve IP addresses and hostnames respectively.

When dealing with files, we temporarily stored the traceroute output in a file and deleted it once its parsing is done, so that we can create a newone for the next value.

Assumptions

1) While using the gethostbyname(), we are considering only the first value of the structure h_addr_list even if it contains multiple values.
2) Starred Hops are those hops that contain only stars.
3) While calculating the response time, we are considering the average of the time values present in the line. And summing the averages across all the lines to get the total response time.
4) If the lastline of the traceroute output contains multiple hops we check for a match with the last hop value.
5) We defined a macro SIZE for the size of the array that holds all the values (IPaddrs, hostnames) in the traceroute output. Since the size of this array is dependent upon

the number of lines in the input files. We might need to change this value wherever necessary.

Usage
./npd

References

1) http://www.traceroute.org
2) http://www.zytek.com/traceroute.man.html
3) www.cplusplus.com