Project 1: Analyzing Internet Network Connectivity

Due: 02/02/2017

Educational Objectives: Refreshing C/C++ programming skills, experiencing text processing techniques, experience using makefile to organize and compile applications.

Statement of Work: Implement a program to study the Internet network connectivity by analyzing BGP routing tables. You may use any C++ features including STL containers and algorithms.

Deliverables: Turn in all the source code files and the makefile in a tar file on blackboard.

Project Description:

The Internet is a collection of large number of networks or autonomous systems (ASes). Each AS owns a number of network prefixes (range of network addresses). Border Gateway Protocol (BGP) is the inter-domain routing protocol used on the Internet for ASes to exchange the reachability information among ASes (or rather, their network prefixes). Each AS has an unique AS number. For example the AS number of the FSU campus network is 2553 and the network prefix owned by the FSU campus network is 128.186.0.0/16. By considering each AS as a node, and the connection between neighboring ASes as a link/edge, we can consider the Internet as an AS-level graph.

In this assignment, you need to develop a program analyze the BGP routing table to get the following connectivity property of the Internet AS graph.: the set of neighboring ASes of each AS. The BGP routing table data trace is provided in the following file (compressed using bzip2).

- BGP routing information base (RIB) data trace
- BGP routing information base (RIB) data trace (the first 10000 lines)

Each line in the file contains a record of the BGP routing information. It follows certain format. The fields in a line are separated by a vertical bar |. Two fields in a line, the 6th and 7th fields, are of particular interests to the Internet inter-domain routing. The 6th field is the network prefix that this record is about. The 7th field is the so called ASPATH, which indicates the sequence of ASes that packets need to traverse to reach the corresponding destination network prefix.. In this project, you only need to work on the 7th field (the ASPATH information). For example, the following line is taken from the data trace file.

TABLE_DUMP|1130191746|B|144.228.241.81|1239|128.186.0.0/16|1239 2914 174 11096 2553|IGP|144.228.241.81|0|-2|1239:321 1239:1000 1239:1011|NAG||

The 6th field is 128.186.0.0/16 (FSU campus network prefix), and the 7th field is 1239 2914 174 11096 2553 (ASPATH). The ASPATH field states that, the rightmost AS (2553, FSU campus network) originates (i.e., owns) the corresponding network prefixes (128.186.0.0/16), and on the way from the leftmost AS (1239) to the rightmost AS (2553), a packet needs to traverse the immediate ASes (2914 174 11096). The ASPATH information expose the neighboring ASes, i.e., the adjacent ASes in the ASPATH are neighbors on the Internet. For example, 1239 and 2914 are neighbors, 2914 and 174 are neighbors, and so are 174 and 11096, 11096 and 2553.

Unfortunately, an ASPATH may have two special cases that you need to pay attention to. First, some of ASPATH contain so called ASSET that we do not know their exact relationship. ASSET is indicated by a pair of square bracket ([]). For example, the following is an ASPATH that contains an ASSET:

1239 1668 10796 [11060 12262]

which was taken from the following line in the data trace:

TABLE_DUMP|1130191716|B|144.228.241.81|1239|24.223.128.0/17|1239 1668 10796 [11060 12262]|IGP|144.228.241.81|0|-2|1239:321 1239:1000 1239:1006|AG|24.95.80.203|

In this example, 11060 and 12262 are in an ASSET. Fortunately, all ASSET occurs at the end (right side) of the ASPATH.

Second, some AS numbers may appear multiple times in an ASPATH, for example, in the following ASPATH, AS number 7911 appears three times, and 30033 appears twice.

1239 7911 7911 7911 30033 30033

which is taken from the following line in the data trace:

TABLE_DUMP|1130191714|B|144.228.241.81|1239|8.3.43.0/24|1239 7911 7911 7911 30033 30033|IGP|144.228.241.81|0|-2|1239:123 1239:1000 1239:1011|NAG||

Fortunately, if an AS number appears multiple times, they appears together (next to each other).

You need to take special care of such lines in the data trace, as discussed below.

Project Requirement:

In this project, you need to develop a program to analyze BGP routing tables to obtain the Internet connectivity information. More specifically, you need to determine the set of neighboring ASes for each AS. When analyzing the neighboring ASes for each AS, you ignore the ASSET part of the ASPATH, but you still need to use the rest of the ASPATH. Moreover, you need to take care of the case that an AS number appears multiple times. *An AS is not the neighbor of its own.*

Output format:

Each output line contains three pieces of information (each piece of information is separated by a space): The AS number of an AS, the total number of neighboring AS of the AS, and the list of the neighboring ASes of the AS. Neighboring ASes in the list are separated by the vertical bar "|" and in increasing order (numerically). The following is an example output line.

2553 3 3506|6912|11096

All the output lines need to be sorted according to the number of neighbors of the ASes in decreasing order. If two ASes have the same number of neighbors, the lines are sorted according to the AS number in increasing order (numerically).

Example run:

$ proj1.x < rib.20051024.2208_144.228.241.81 > ASdegree.txt

ASdegree.txt is also given so that you can compare with your results. (for the smaller data set, here is the output ASdegree_10000.txt

Provided files

- **proj1.x**: executable compiled on linprog.
- **BGP routing information base (RIB) data trace**
- **BGP routing information base (RIB) data trace (the first 10000 lines)**
- **ASdegree.txt**
- **ASdegree_10000.txt**

Some help information

Some brief introduction of BGP is available at here or here. Or you can search online.

Information regarding how BGP data trace is collected is available at the route view project site.