About the program

What it does

Processing of datasets provided by the RIPE Atlas Internet measurement network. Downloads 1 hour of traceroute date and creates summaries of possible outages. It also provides the topologies of the recorded routes and some other cool things.

Release note

14.05.21

Unique routes are now identified by combining the probe ID of the source with the IP of the destination, i. e. "6743-to-45.77.229-242", rather than by combining the source IP and destination IP. The reason we no longer want to identify the source by its IP is because multiple probes will often share the same public IP. In rare cases a probe ID may still get a new public IP. In the rare cases when multiple source IPs are tied to a single probe, all IPs used are listed in chronological order as "src_addr".

07.05.21

Only outages from routes with at least one recorded successful traceroute are now considered. This drastically reduces the output of the program, making it more manageable. Have also revamped the topologies to include more data and RTT information. Additionally there are many minor fixes and improvements. Some unused code has also been removed and a few comments have been added.

15.04.21

The current version of the program has a looser definition of an outage than previously which increases the amount of summaries by around 20 times from before. This may or may not be kept in possible future revisions of the program.

Install and use

Running atlas-gap-ana.py

1. Have Python3 installed. You may also want to have some kind of package manager like Pip.
2. Clone this repository to a folder on your server/computer.
3. Install the dependencies from "requirements.txt". If you are using Pip, you can run `pip install -r requirements.txt`.
4. To launch the program, use the pattern `python atlas-gap-ana.py 2021-04-10T0100` to specify the hour you want analyzed.

How to use the program

It's intended to be run 24 times in parallel on a powerful server. A sample Unix script is provided which runs the program once for every hour on your chosen date. To run the script, use the pattern `./atlas-gap-ana`

`2021-04-10`.

Explanation of logs

The program processes huge collections of traceroutes provided by the RIPE Atlas Internet measurement network. The original logs may be found here. Documentation for the logs are found here. The program in this repository turns these traceroute logs into 3 new types of logs.

Log summary

Describes the full log that was analyzed. One such log pr. hour.

Attribute	Description
total_traceroutes	Total amount of traceroutes in original log
event_type	Type of log, i. e. "Log summary".
start_time	Earliest timestamp from original log.
end_time	Latest timestamp from original log.
total_succeeded	Total successful traceroutes in original log.
total_failed	Total failed traceroutes in original log.
failed_at_probe	Amount of traceroutes that failed before the first hop.
total_probes	Total unique probes recorded in original log.
version	Version of atlas-gap-ana.py

Summary

Describes unique routes with at least one outage recorded.

Attribute	Description
pair_id	Unique combination of "from" and "to" IP.
event_type	Type of log, i. e. "Summary".
dst_addr	Destination IP.
src_addr	"From" IP. In the rare case when a probe changes its IP, the IPs for each traceroute are listed chronologically.
prb_id	The ID of the probe performing the traceroutes tied to this unique route.
start_time	Timestamp of earliest traceroute on the route.
last_response	Timestamp of final traceroute on the route.
traceroutes_recorded	Total amount of traceroutes recorded on the route.

Attribute	Description
failed_traceroutes	Percentage of traceroutes recorded on the route that failed to reach their destination.
outage_count	Amount of outages found on the route.
outage_at_log_start	Whether the route is having an outage at the start of the hour.
outage_at_log_end	Whether the route is having an outage when the hour ends.
incomplete_outages	Amount of traceroutes with not enough data to determine causes.
total_dupes	Amount of dups found in last set of ICMP replies.
total_late	Amount of late replies found in the last set of ICMP replies.
total_outage_errors	All errors from all outages on this route.
total_outage_duration_in_seconds	Total sum of outage time on this route.
max_hop	Length of longest path from source to destination seen in a traceroute.
min_hop	Length of shortest path from source to destination seen in a traceroute.
topology	Topology of the given route.

Outage

Describes each unique probable outage.

Attribute	Description
pair_id	Unique combination of "from" and "to" IP.
event_type	Type of log, i. e. "Outage".
order	Chronological order of outage pr. unique route.
dst_addr	Destination IP.
src_addr	"From" IP. In the rare case when a probe changes its IP, the IPs for each traceroute are listed chronologically.
prb_id	The ID of the probe performing the traceroutes tied to this outage.
start_time	Timestamp of earliest traceroute in the outage.
last_response	Timestamp of last traceroute in the outage.
outage_at_log_start	Whether the particular outage is going when the hour starts.
outage_at_log_end	Whether the particular outage is going when the hour ends.
failure_count	Amount of failed traceroutes the outage is based on.

Attribute	Description
polling_frequency	Average amount of seconds between each traceroute. Set to 0 if unknown.
duration	Length of outage. Set to 0 if unknown.
last_reply_from	Last IP to to reply to failed traceroutes.
AS_number	AS number last_reply_from belongs to.
AS_owner	Owner of AS last_reply_from belongs to.
location	Country and maybe city where last_reply_from is located.
coordinates	Map coordinates of last_reply_from's location.
outage_errors	Error messages from the outage.
error_type	Program's best guess at main cause of outage.
max_hop	Length of longest failed path for a traceroute.
min_hop	Length of shortest failed path for a traceroute.
topology	List of failed routes used by the failed traceroutes.