# Loon Network Data

## Authors

Frank Uyeda, Filip Zivkovic, Marc Alvidrez, Aswin Alexander
Loon LLC

## Introduction

Loon sought to provide data connectivity to under-served regions via 4G LTE cell service hosted on free-floating, high-altitude balloons. To backhaul data, the balloons used point-to-point radio links to connect with each other and to transceivers connected to the network core on the ground (aka. "ground stations"). These links were selected & commanded by a centralized Temporo-Spatial Software Defined Network (TS-SDN) controller running on the ground. More details about the system architecture and operation can be found in the Loon Library or in "*SDN in the Stratosphere: Loon's Aerospace Mesh Network*" from the proceedings of SIGCOMM 2022.

The dataset provided here consists of internal state from the TS-SDN and network telemetry gathered from serving commercial traffic and R&D experiments. Loon's networking use case is unique in several ways:

- Balloons and their base stations/eNodeBs were located in the stratosphere and changed position as they were pushed by the wind.
- The network was comprised of a combination of LTE, WiFi, E-band, Wired technologies
- The topology of our E-band backhaul meshes spanned thousands of kilometers and changed frequently.
- eNodeBs entered and exited the cellular network depending on their position and the availability of power and backhaul connectivity.

Loon collected data using system telemetry from Loon assets (balloons, ground stations, etc) in the field. These logs were extracted from various storage systems and processed for external publication.

## About the data

The data is presented in a tabular format, where each table is stored as a comma-separated values ASCII file (i.e. CSV format). Each file is compressed using bz2 (bzip2, a block-sorting file compressor.  Version 1.0.8, 13-Jul-2019). There are 5 tables in total:

- Network connectivity probes, in backhaul.csv.bz2, contains result of network reachability probes from the ground to nodes within the network via different layers of the network's control and data planes at points in time
- A Link Intent is created by the TS-SDN to indicate its desire for a link between two node's interfaces, and to track the state of the link over time. Link intents over time, in link_intents.csv.bz2, contains a change log of state transitions of each attempted link.
- The eNodeB is the 4G component that manages sector antennas of the LTE base station. eNodeB stats, in enodebstats.csv.bz2, contains data service download and upload statistics provided to users per eNodeB.
- Flight regions, in flight_regions.csv, contains geographic region locations for each flight at a given time.
- To select the links that it wishes to establish or withdraw, the TS-SDN generated Transceiver Link Reports for pairs of transceivers. Transceiver Link Reports record the forecasted radio link performance and the sources of attenuation for a given set of transceiver parameters for a time in the future. Forecasted RF performance incorporates the expected spatial geometry of the nodes at the forecast time, and the forecasted weather along the transmission vector. Transceiver Link Reports are stored in per-hour-shared files of the format link_reports-*.csv.bz2

### Naming Conventions

Network Nodes hosted on balloons are typically represented by the flight's name. Flight names have the format "${flight_prefix}-${flight_number}" , where flight_prefix is 1 or 2 upper-case characters and flight_number is a zero-padded 3-digit integer (e.g. "LN-127").

Ground stations used either WiFi or e-Band transceivers. Ground Stations with e-Band transceivers have names with the format "leg${antenna_number}.${location_code}${site_number}", where location_code is a 3 letter string, and site_number & antenna_number are zero-padded 2-digit ordinal integers (e.g. "leg01.nbo01"). Ground stations with WiFi transceivers have the format "lwg${antenna_number}.${location_code}${site_number}" (e.g. "lwg01.wmc01").

Network Nodes that could not be mapped to human readable names are represented by unique identifiers. Unnamed ground stations have ids of the form "gs/${id}" (e.g. "gs/698157"). All other unnamed network nodes have ids such as "e17774cb-5799-41d0-a069-80312c0233af".

## Network connectivity probes

Network connectivity probes table has columns as per the table below:

| Column Name | Column Description |
|---|---|
| timestamp | human readable time (i.e. "2020-04-18 09:30:00") |

| | |
|---|---|
| node | The identifier for a Loon flight (e.g. "P-646") |
| estimated_backhaulable | whether our backend system believed that a topology solution existed that would allow this node to connect to a ground station reliably, either directly or via the mesh; “true”, “false”, or empty string; empty string means that backhaulability could not be estimated at that time |
| batman_probe | whether node can be reached by utilizing B.A.T.M.A.N protocol that is running on our mesh at the time; “true”, “false”, or empty string; empty string means that measurement could not be obtained at that moment |
| inband_cdpi | whether we can issue commands to the node by using our inband control plane connection (from a ground station, via mesh, to the flight); “true”, “false” or empty string; ; empty string means that measurement could not be obtained at that moment |
| sdn_probe | whether we can reach node via the data plane path configured by our software-defined networking commands; “true”, “false” or empty string; ; empty string means that measurement could not be obtained at that moment |

## Link intent change log

The Link Intents table has columns as per the table below:

| **Column Name** | **Column Description** |
|---|---|
| timestamp | human readable time (i.e. "2018-11-26 08:00:02.153722") of the creation or update to this link intent record |
| identifier | unique identifier of the intent (e.g. "Temporospatial Topology & Routing:Intent:-9223372036843806981") |
| node_a_identifier | The identifier for a Loon flight or ground station for which the intent was issued; when possible, name of the node is used (e.g. LN-158) and otherwise identifier is used (e.g. "gs/698157", or "e17774cb-5799-41d0-a069-80312c0233af") |
| node_b_identifier | |
| node_a_interface | identifier of interface on a node for which the intent was issued (“wifi0”, “mmwave0”, “mmwave1” or “mmwave2”) |
| node_b_interface | |
| state | The new or updated state of the link intent; values can be:<br>- “INSTALL_REQ”, when link is requested to be established in the future but the process of doing so has not started yet |

| | |
|---|---|
| | - “INSTALLING”, when the process of establishing the link has begun but has not completed yet<br>- “INSTALLED”, when the process of establishing the link has completed and the link is ready for use;<br>- “FAILED”, when link cannot be established or maintained despite our system’s instructions to do so<br>- “WITHDRAW_REQ”, when a link is requested to be removed but the process of doing so has not started yet.<br>- “WITHDRAWING”, when process of removing the link has begun but has not completed<br>- “WITHDRAWN”, when process of removing the link has completed |

## eNodeB stats

The eNodeB stats table has columns as per the table below:

| **Column Name** | **Column Description** |
|---|---|
| timestamp | human readable time (i.e. "2020-04-18 09:30:00") |
| enodeb | unique identifier of enodeb; when available, node name is used; if there is more than one sector per node then node name is followed by board number and sector id; (e.g. "0-264" or "LN-316/1/0") |
| pdcp_download | average download/upload throughput, in bytes per second through P-gateway, for the enodeb in question since the last timestamp that appears in the file |
| pdcp_upload | |

## Flight regions

The Flight Regions table has columns as per the table below:

| **Column Name** | **Column Description** |
|---|---|
| timestamp | human readable time (i.e. "2020-04-18 09:30:00") |
| node | The identifier for a Loon flight (e.g. "LN-100") |
| region | the name of the region the flight is in, or empty string (when flight is out of any region); regions present are “Baas Test”, “Loreto Test Contiguous”, “Kenya All” and “Mozambique”. |

## Transceiver link reports

The Transceiver Link Reports table has columns as per the table below:

| Column Name | Column Description |
|---|---|
| record_timestamp | The date and time at which this row was recorded. Times are at second granularity in UTC timezone.<br>(e.g "2020-12-05 22:15:10") |
| node_a_indentifier | The identifier for a Loon flight or ground station on one end of a potential link. |
| node_b_indentifier | Same as 'node_a_identifier' |
| node_a_interface | The identifier for both the physical transceiver and the logical network interface attached to a node (whether a flight or ground station) on one end of a potential link.<br>(e.g. "mmwave0") |
| node_b_interface | Same as 'node_a_interface' |
| band_profile_id | The radio frequency band and channel width to be analyzed. Multiple 420 Mhz channels were available for use in both the 70 Ghz and 80 Ghz bands. The dataset has been culled to only include the channel in each band with the largest link_margin_db.<br>(e.g. "70G-420w" or "80G-420w") |
| forecasted_timestamp | The future date and time at which we are forecasting the spatial geometry and RF analysis for this potential link. Times are at second granularity in UTC timezone.<br>(e.g "2020-12-05 22:15:10") |
| range_m | The line-of-sight distance in meters between the identified transceivers on the identified nodes.<br>(e.g. "74631.0") |
| propagation_delay_usec | The modeled radio signal propagation delay in microseconds between the identified transceivers on the identified nodes.<br>(e.g. "248") |
| channel_center_freq_Ghz | The center frequency in Ghz of the channel with the largest link_margin_db within the band. Each channel is 420 Mhz wide.<br>(e.g. "71.5") |
| transmit_power_watts | The maximum available transmit power (in Watts) that our transceiver would be able to generate at the specified center frequency.<br>(e.g. "0.236") |
| effective_isotropic_radiated_power_dbw | The effective isotropic radiated power in dBw which would be radiated in the direction of the receiving transceiver when |

| | |
|---|---|
| | transmitting with 'transmit_power_watts'. The Loon antennas supplied a ~44dB gain along the beam axis.<br>(e.g. "37.73") |
| free_space_loss_db | The expected free space attenuation between the transmitter and receiver when using the indicated channel and the forecasted range.<br>(e.g. "-166.99") |
| atmospheric_loss_db | The expected atmospheric attenuation between the transmitter and receiver when using the indicated channel, the forecasted transmission vector, and the traversed altitude range.<br>(e.g. "-0.1") |
| rain_fade_loss_db | The expected attenuation due to rain between the transmitter and receiver when using the indicated channel, and predicted rain rates along the forecasted transmission vector.<br>(e.g. "0.0") |
| cloud_fog_loss_db | The expected attenuation due to atmospheric water content between the transmitter and receiver when using the indicated channel, and the predicted cloud/fog density along the forecasted transmission vector.<br>(e.g. "0.0") |
| cummulative_propagation_loss_db | The sum of all modeled sources of RF attenuation.<br>(e.g. "-167.09") |
| received_isotropic_power_dbw | The effective isotropic radiated power in dBw which would be incident upon the receiving transceiver from the direction of the transmitter. Should be equivalent to effective_isotropic_radiated_power_dbw + cummulative_propagation_loss_db<br>(e.g. "-129.36") |
| power_at_receiver_output_dbw | The expected power received by the target transceiver. The Loon antennas supplied a ~44dB gain along the beam axis.<br>(e.g. "-86.96") |
| data_rate_Mbps | The highest data rate supported by the transceiver which will result in a link margin above 5 dB. If none exists, then the lowest supported data rate, so long as the link margin is above -5 dB.<br>(e.g. "987.0") |
| link_margin_db | The expected link margin in decibels available when transmitting on the indicated channel at the indicated power and data rate.<br>(e.g. "29.44") |

| | |
|---|---|
| accessibility | An indication of whether accessibility is expected to exist for the modeled link at the indicated data rate. We say accessibility "exists" for links with a link margin >5 dB, and is "marginal" for links with +/- 5 dB of margin. Links with less than -5 dB link margin are considered "inaccessible" and excluded from the data set.<br>(e.g. "EXISTS" or "MARGINAL") |