

Simulated datasets

- [Simulated datasets](#)
- [Description](#)
 - [Telecom Italia dataset](#)
 - [Where](#)
 - [How](#)
 - [Details](#)
 - [Other references](#)
 - [Viavi dataset](#)
 - [Where](#)
 - [How](#)
 - [Details](#)

Simulated datasets

Description

For AI/ML in general, and In the context of Energy Savings in particular, it is important that datasets offering training data exist. The SIM project proposes currently 2 data sets to be leveraged: one based on real network data, coming from Telecom Italia, and another one based on synthetic data created with the help of Viavi.

Telecom Italia dataset

Where

The TIM data set representing the “Internet” usage from the network, for 2 months (Nov and Dec 2013, timestamps every 10 minutes) is published in O-RAN SC nexus:

- https://nexus3.o-ran-sc.org/repository/datasets/sms-call-internet-mi-2013-11-01_parsed.tar.gz
- https://nexus3.o-ran-sc.org/repository/datasets/sms-call-internet-mi-2013-11-02_parsed.tar.gz
- ...
- https://nexus3.o-ran-sc.org/repository/datasets/sms-call-internet-mi-2014-01-01_parsed.tar.gz

How

Just download the data set and extract.

Details

The data is parsed, meaning we pruned the voice and SMS usage, keeping only the Internet traffic. The values themselves for the usage are normalised to a value known only to TIM (assuring some anonymisation).

It contains data for November and December 2023, with timestamps every 10 minutes. We have a different file for each of the days.

The data is from the city of Milano. As described in the references, the city was divided into 10.000 squares (roughly a grid of 100 x 100 squares). That is the *SquareID* (between 1 and 10000) column. The timestamp is the start time interval of the measurement. The end time will be start time + 10 minutes. Please note that the timestamp is GMT+1, Milano local time. The InternetTraffic represents “number of CDRs generated inside a given Square id during a given Time interval.”. It does not have any unit of measure, because it is somehow normalised such that it is anonymised. It has no meaning in itself, but we can see patterns of how that value increases and decreases over time.

Other references

- <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/EGZHFV>
- <https://www.nature.com/articles/sdata201555>

Viavi dataset

Where

The two files of this dataset are stored here:

- <https://nexus3.o-ran-sc.org/repository/datasets/CellReports.csv>
- <https://nexus3.o-ran-sc.org/repository/datasets/UEReports-flow.csv>

How

Download the data set. Files will be password protected. Whoever is interested in using the data can comment on this page and will receive the file password via email.

Details

The dataset contain synthetic data generated by the simulator, consisting of two files: Cell metrics and UE metrics. Each file contains various metrics specific to either cells or user equipment (UE). Data collection occurred over a 7-day period, from January 1, 2023, to January 7, 2023. The 4G cells use frequencies B2 and B17, while the 5G cells use frequency N77, distributed randomly within a 10 square kilometer simulation area, utilizing the default scenario settings in the simulator.

The data is timestamped in Unix format.

Cell Metric

Parameter	Description	Unit
DRB.UEThpDI	Downlink throughput	Gbps
DRB.UEThpUI	Uplink throughput	Gbps
RRU.PrbUsedDI	Downlink Physical Resource Blocks (PRBs) used	N/A
RRU.PrbUsedUI	Uplink Physical Resource Blocks (PRBs) used	N/A
RRU.PrbAvailDI	Number of Physical Resource Blocks (PRBs) available for downlink	N/A
RRU.PrbAvailUI	Number of Physical Resource Blocks (PRBs) available for uplink	N/A
RRU.PrbTotUI	Total usage of Physical Resource Blocks (PRBs) on the uplink	
RRU.PrbTotDI	Total usage of Physical Resource Blocks (PRBs) on the downlink	
RRC.ConnMean	Mean number of UEs in RRC connected mode	N/A
RRC.ConnMax	Maximum number of UEs in RRC connected mode	N/A
QosFlow.TotPdcPduVolumeUI	Uplink data volume (PDCP PDU) delivered from gNB-DU to gNB-CU	Mbits
QosFlow.TotPdcPduVolumeDI	Downlink data volume (PDCP PDU) delivered from gNB-CU to gNB-DU	Mbits
PEE.AvgPower	Average power utilization	watts (W)
PEE.Energy	Energy utilization	kilowatt-hours (khW)
RRU.MaxLayerDIMimo	Average maximum scheduled layer number under MIMO scenario in DL	N/A
CARR.AverageLayersDI	Average value of scheduled MIMO layers per PRB on the DL	N/A
RRC.ConnEstabAtt.mo-Data	Number of UE RRC connections to the cell by "mobile oriented data" cause	N/A
RRC.ConnEstabAtt.mo-VoiceCall	Number of UE RRC connections to the cell by "mobile oriented voice call" cause	N/A
RRC.ConnEstabAtt.mo-VideoCall	Number of UE RRC connections to the cell by "mobile oriented video call" cause	N/A
RRC.ConnEstabSucc.mo-Data	Number of successful UE RRC connections to the cell by "mobile oriented data" cause	N/A
RRC.ConnEstabSucc.mo-VoiceCall	Number of successful UE RRC connections to the cell by "mobile oriented voice call" cause	N/A
RRC.ConnEstabSucc.mo-VideoCall	Number of successful UE RRC connections to the cell by "mobile oriented video call" cause	N/A
RRC.ConnEstabFailCause.NetworkReject	Number of unsuccessful UE RRC connections to the cell rejected by the network	N/A

UE Metric

Parameter	Description	Unit
RRU.PrbUsedUI	Mean uplink Physical Resource Blocks (PRBs)	N/A
RRU.PrbUsedDI	Mean downlink Physical Resource Blocks (PRBs)	N/A
DRB.UEThpUI	Uplink throughput	Gbps

DRB.UEThpDI	Downlink throughput	Gbps
TB.TotNbrUI	Total number of uplink Transport Blocks (TBs)	N/A
TB.TotNbrDI	Total number of downlink Transport Blocks (TBs)	N/A
DRB.UECqiUI	UE's uplink CQI	N/A
DRB.UECqiDI	UE's downlink CQI	N/A