Characterizing and Modeling Mobile Networks User Traffic at Millisecond Level

Pablo Fernández Pérez, Claudio Fiandrino, Joerg Widmer

ACM WiNTECH 2023

Developing the Science of Networks

Introduction

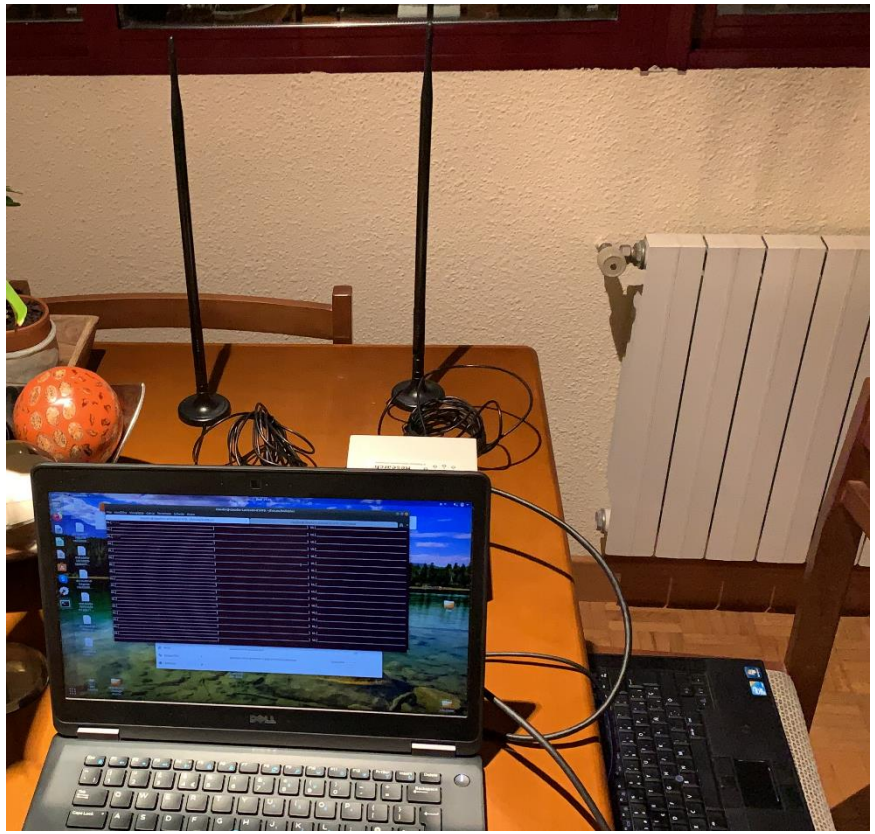
- Mobile traffic data often unavailable due to various reasons
- Lack of mobile datasets limits research advances
- We fill this gap making available a large dataset

Year	Operator	Zone	Carrier Frequency (MHz)	ID	Total Time (h)	Total Users	Traffic Volume (GB)	Dataset Size (GB)
2020	A	I	796	BS1	247	72 321 636	388 691	11
2020	B	I	1815	BS2	227	23 508 594	242 406	3
2020	A	I	2650	BS3	74	10 493 540	18 403	0.24
2021	B	II	816	BS4	163	14 813 731	579 376	12
2021	C	II	1835	BS5	313	59 463 421	2 378 256	27
2021	A	II	2650	BS6	353	33 650 085	1 308 064	11

Motivation and Contribution

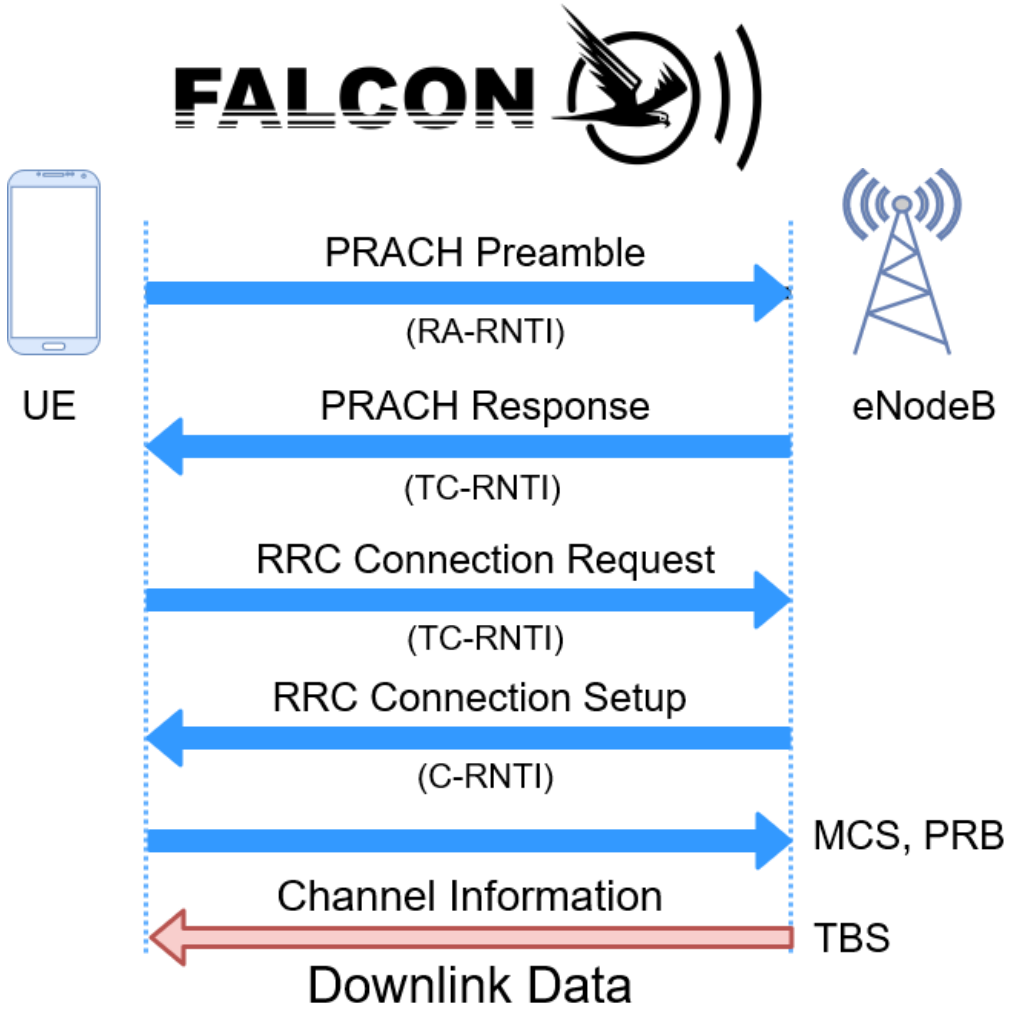
- In-depth characterization of millisecond-level mobile traffic
- We make open:
  - Large dataset from multiple Base Stations (BSs)
  - Real LTE control traffic data at millisecond level
  - Valuable for research (especially data-driven/AI driven optimizations)

At COVID time!



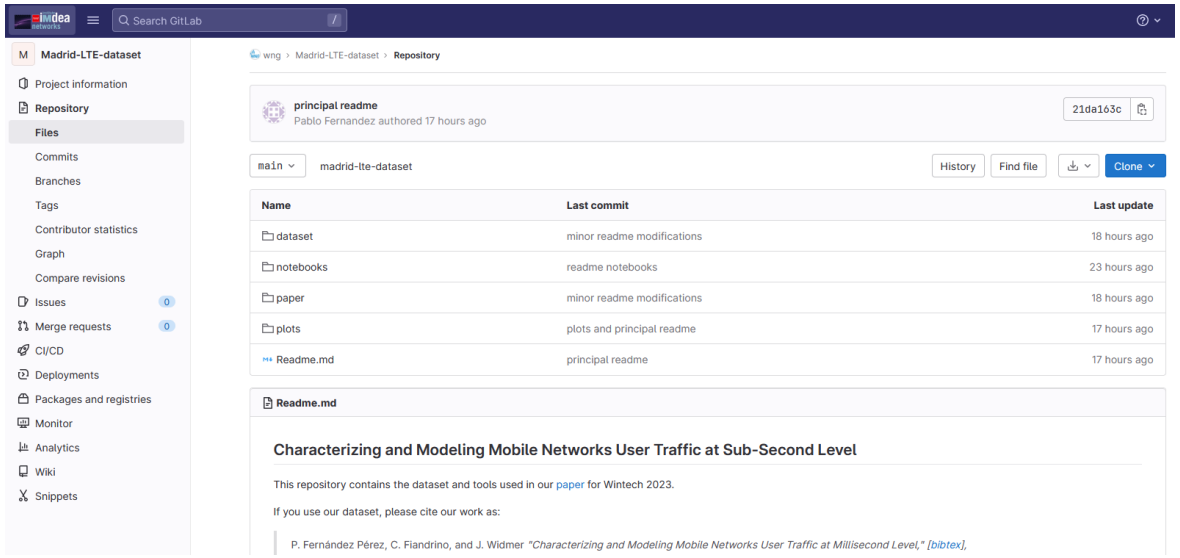
The Decoded Data

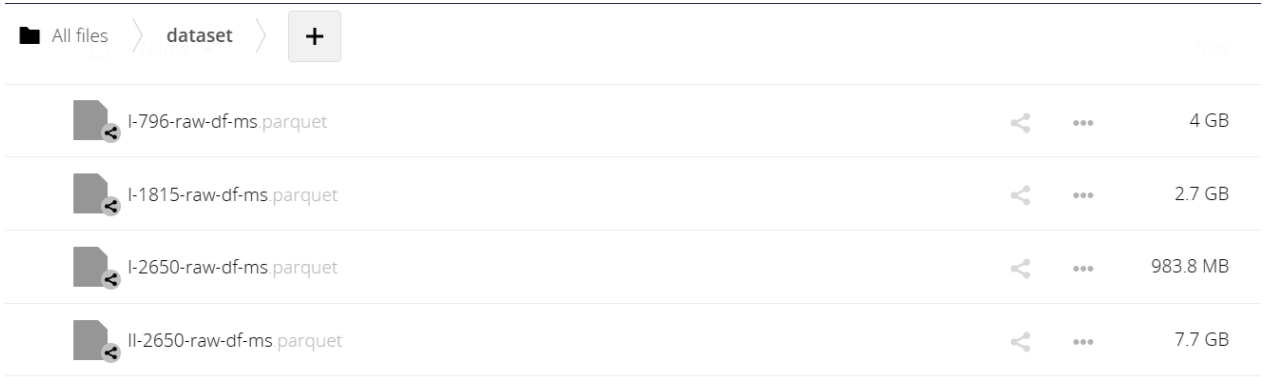
- **Information:**
  - Temporary user ID (RNTI) associated with the user
  - Frame ID containing traffic allocation for each C-RNTI
  - Associated transport block size (TBS)
  - Transmission details: Modulation and Coding Scheme (MCS), and utilized Physical Resource Blocks (PRB)
- **User identification:**
  - Analyze inter-transmission times between identical RNTIs.
  - Set a time threshold of 10 s for user lifetime.



Our Dataset in a Nutshell

- Repository: data and code
- Raw dataset 1 ms granularity
- Minimal processing (RNTI scrambling)
- Also stored 1 s granularity processed data

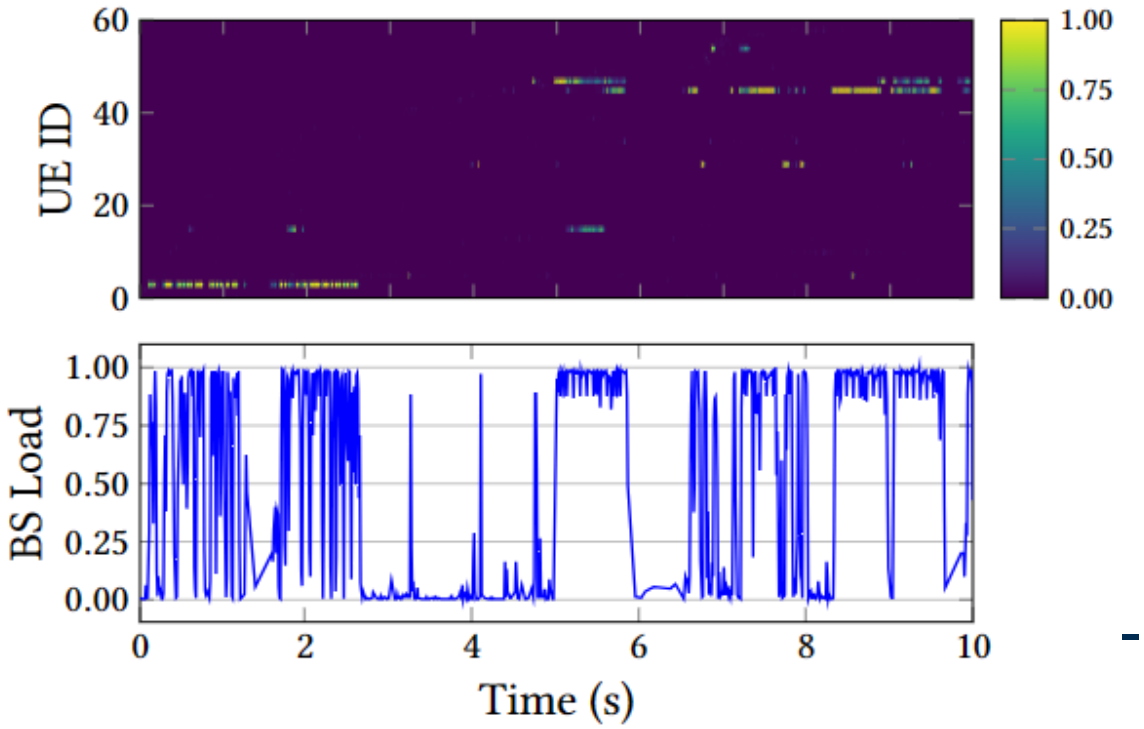


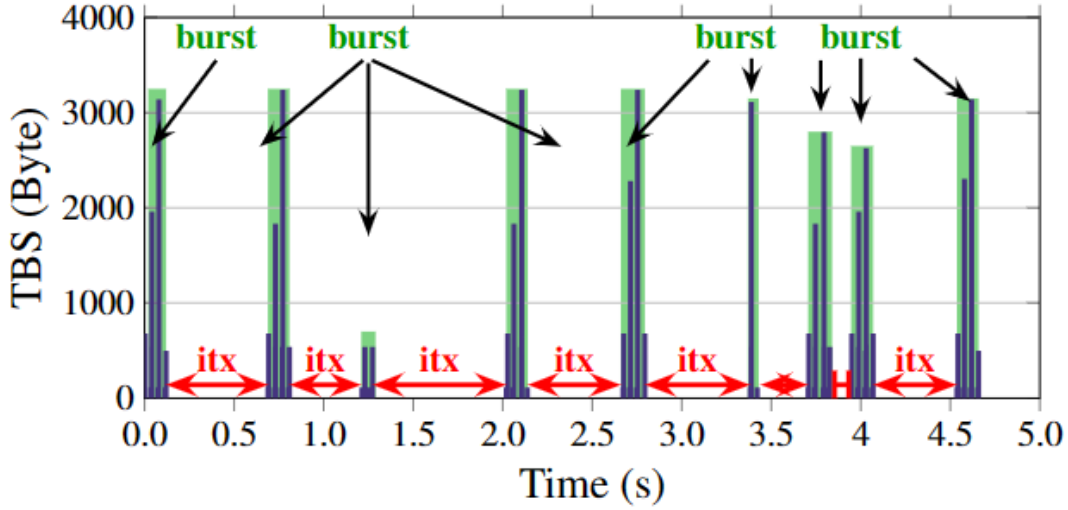


https://git2.networks.imdea.org/wng/madrid-lte-dataset

Terminology

Looking traffic at BS level and at user level





- Burst: consecutive non-zero TBs
- Intertransmission times (itx): time elapsed between consecutives TBs

Results overview

Traffic at BS level

- User-Traffic distribution.
- Self similarity and long-range dependence.

Traffic at user level

- 2d-histograms of:
  - tbs,nº occurences, nº users.
  - itx,nº occurences, nº users.

Number of RRC connected users

- Bimodal distribution.

Traffic at BS level

Traffic at BS level

- User-Traffic distribution.
- Self similarity and long-range dependence.

Traffic at user level

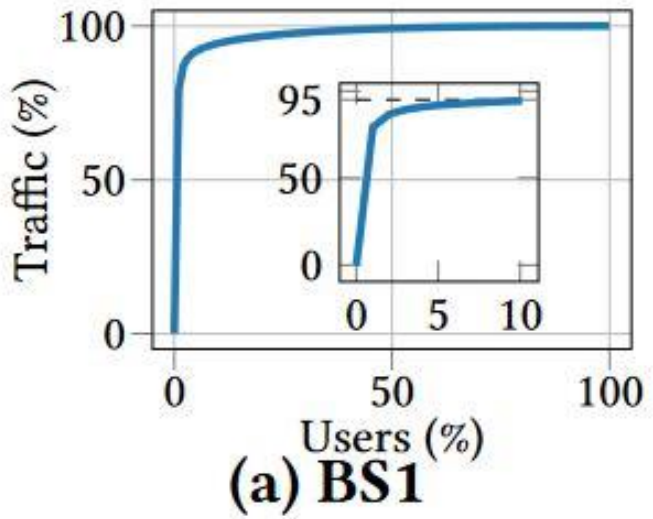
- 2d-histograms of:
  - tbs,nº occurences, nº users.
  - itx,nº occurences, nº users.

Number of RRC connected users

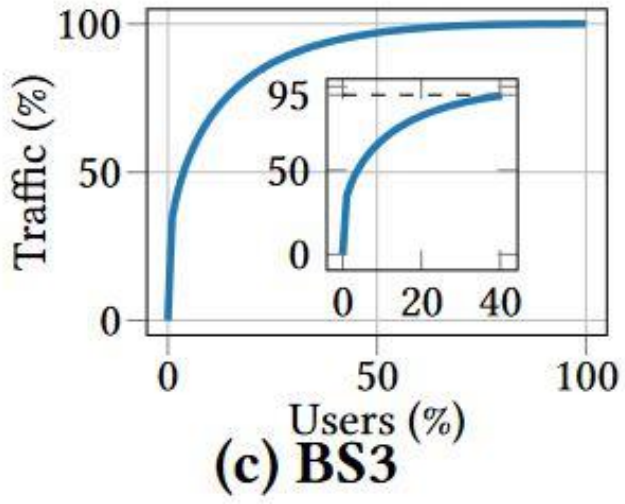
- Bimodal distribution.

Traffic at BS level: User distribution

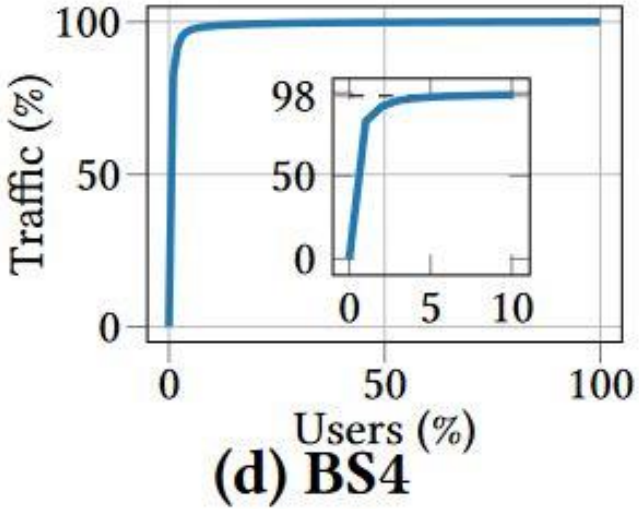
- Top 10% of the users consume 90% of the traffic
- Only BS3 has a slightly different distribution where top 30% of the users cosume 90% of the traffic



(a) BS1



(c) BS3

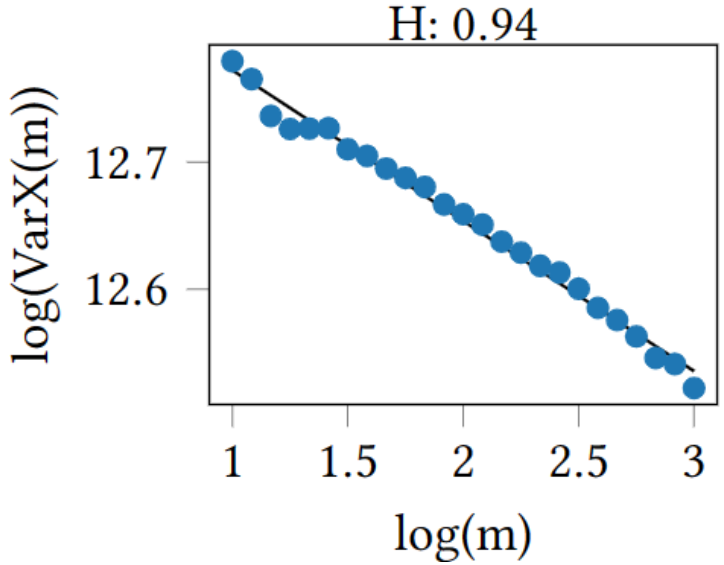


(d) BS4

Traffic at BS level: Hurst Parameter

- Long range dependence (LRD):

the sum of the complete sequence of the autocorrelation function is infinite



- Self-similarity: statistical similarity across different scales (Hurst parameter *H*)

$$x_{at} \sim a^H x_t$$

- Self-similarity more evident for downlink than uplink traffic

Traffic at User Level

Traffic at bs level

- User-Traffic distribution.
- Self similarity and long-range dependence.

Traffic at user level

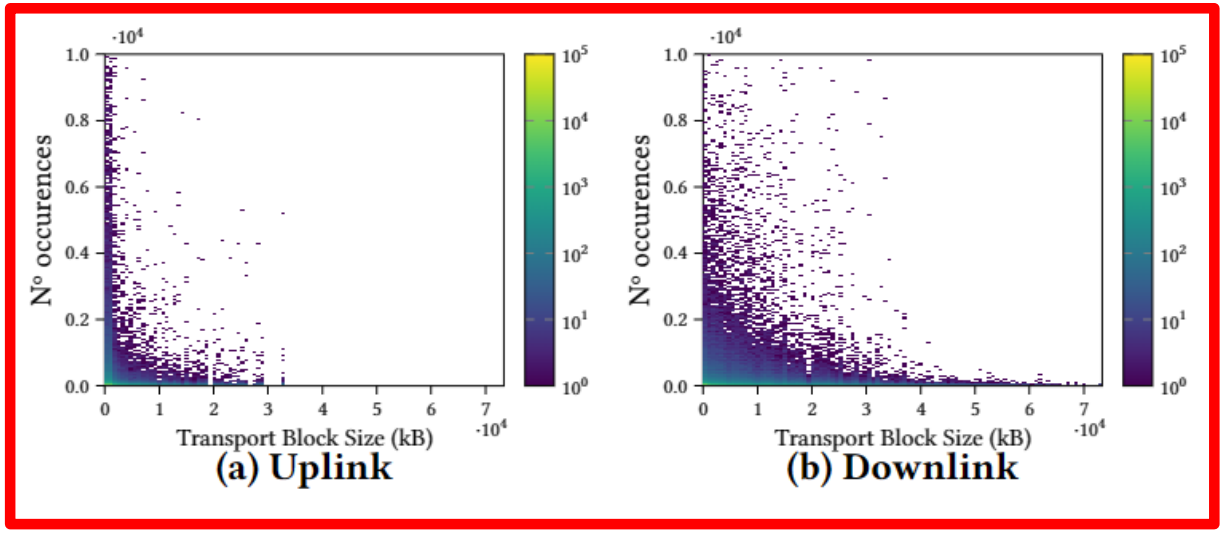
- 2d-histograms of:
  - tbs,nº occurences, nº users.
  - itx,nº occurences, nº users.

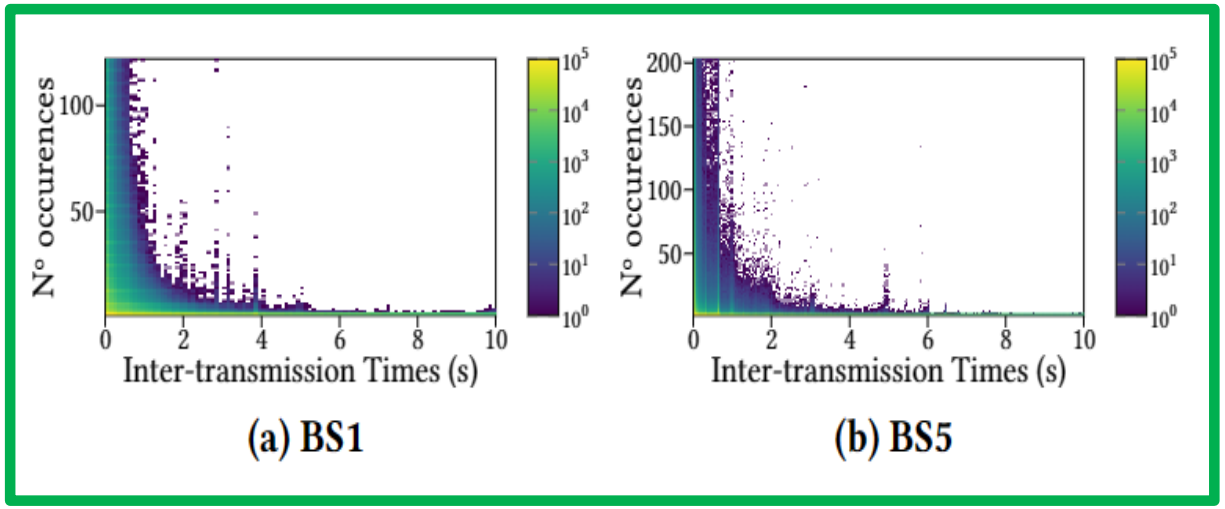
Number of RRC connected users

- Bimodal distribution.

Analysis of User Level Traffic

- Transport block size and inter-transmission times (itx) analysis.
- Separate analysis from uplink and downlink traffic.
- Differences are attributed to traffic nature and resource allocation policies





Number of RRC connected users

Traffic at bs level

- User-Traffic distribution.
- Self similarity and long-range dependence.

Traffic at user level

- 2d-histograms of:
  - tbs,nº occurences, nº users.
  - itx,nº occurences, nº users.

Number of RRC connected users

- Bimodal distribution.

Number of RRC Connected Users

- These users are potentially active (in either uplink or downlink)
- Number of RRC connected users follows a bimodal distribution
- We find all BSs following a bimodal distribution with significant shape differences

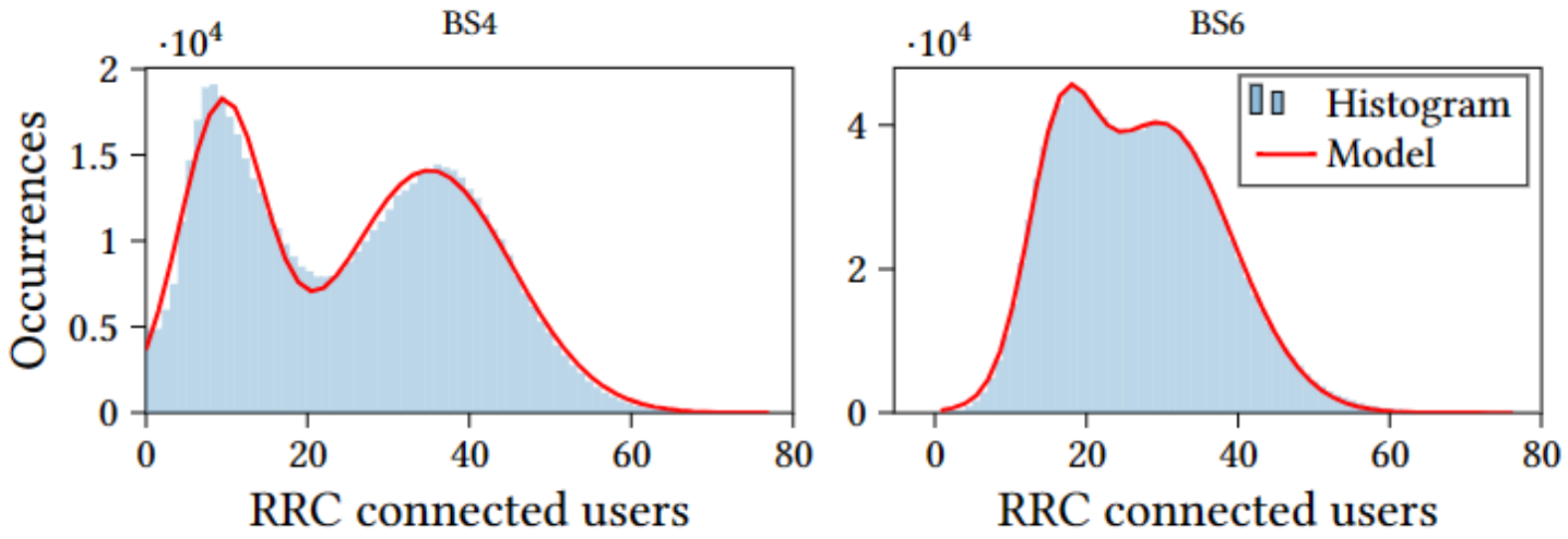


Figure 10: Bi-modal distribution

Take-Home Messages

- Large real-world LTE traffic dataset at ms level granularity
- Enables data-driven research as well as AI-based network optimization
- Important for research reproducibility

https://git2.networks.imdea.org/wng/madrid-lte-dataset