Characterizing and Modeling Mobile Networks User Traffic at Millisecond Level

Pablo Fernández Pérez, Claudio Fiandrino, Joerg Widmer

ACM WiNTECH 2023
• 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
• 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)
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
Looking traffic at BS level and at user level

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

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

Traffic at BS level

Traffic at user level

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

- Bimodal distribution.

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.
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 consume 90% of the traffic
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 $H$).

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,$n^o$ occurences, $n^o$ users.
  - itx,$n^o$ occurences, $n^o$ 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 occurances, nº users.
  - itx, nº occurances, nº users.

Number of RRC connected users
- Bimodal distribution.
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