Spatio-temporal analysis of mobile Internet traffics

Anh-Dung Nguyen, Razvan Stanica, Marco Fiore

Urbanet team, Inria – CITI Lab, Insa Lyon

Outline

Context
Dataset
Visualization & analysis
Simpson paradox
Conclusion

Context

- Exponential growth of mobile Internet
    - 2.5 exabytes/month (2014)
    - Tenfold increase by 2020

- Increased diversity of mobile applications

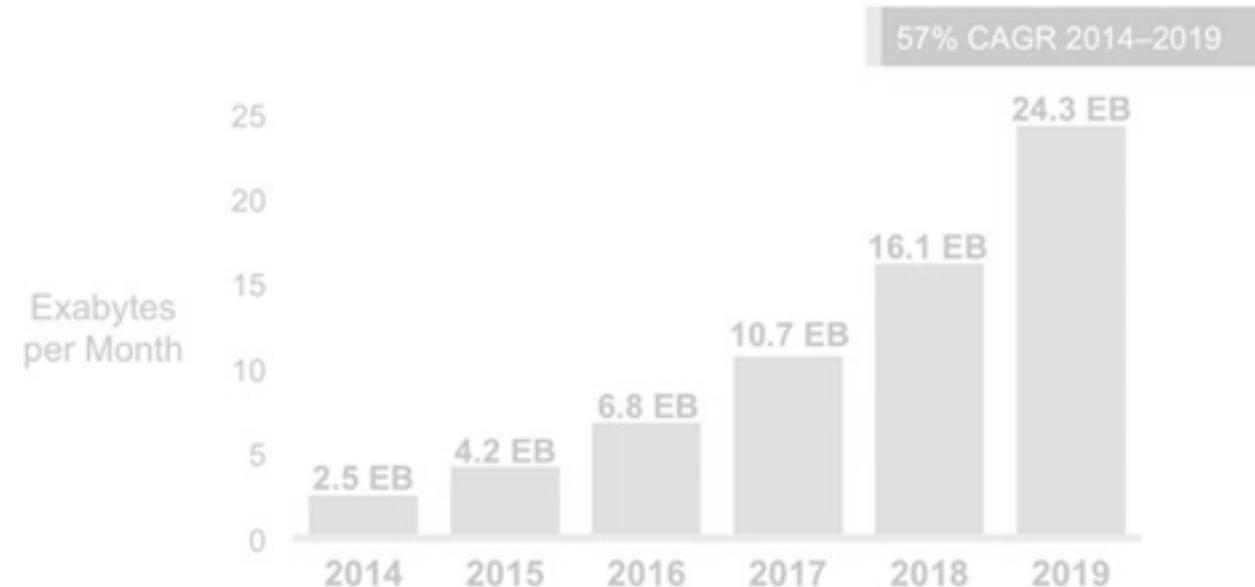- Increased diversity of mobile devices

- The development of IoT

Technical and economical challenges for operators
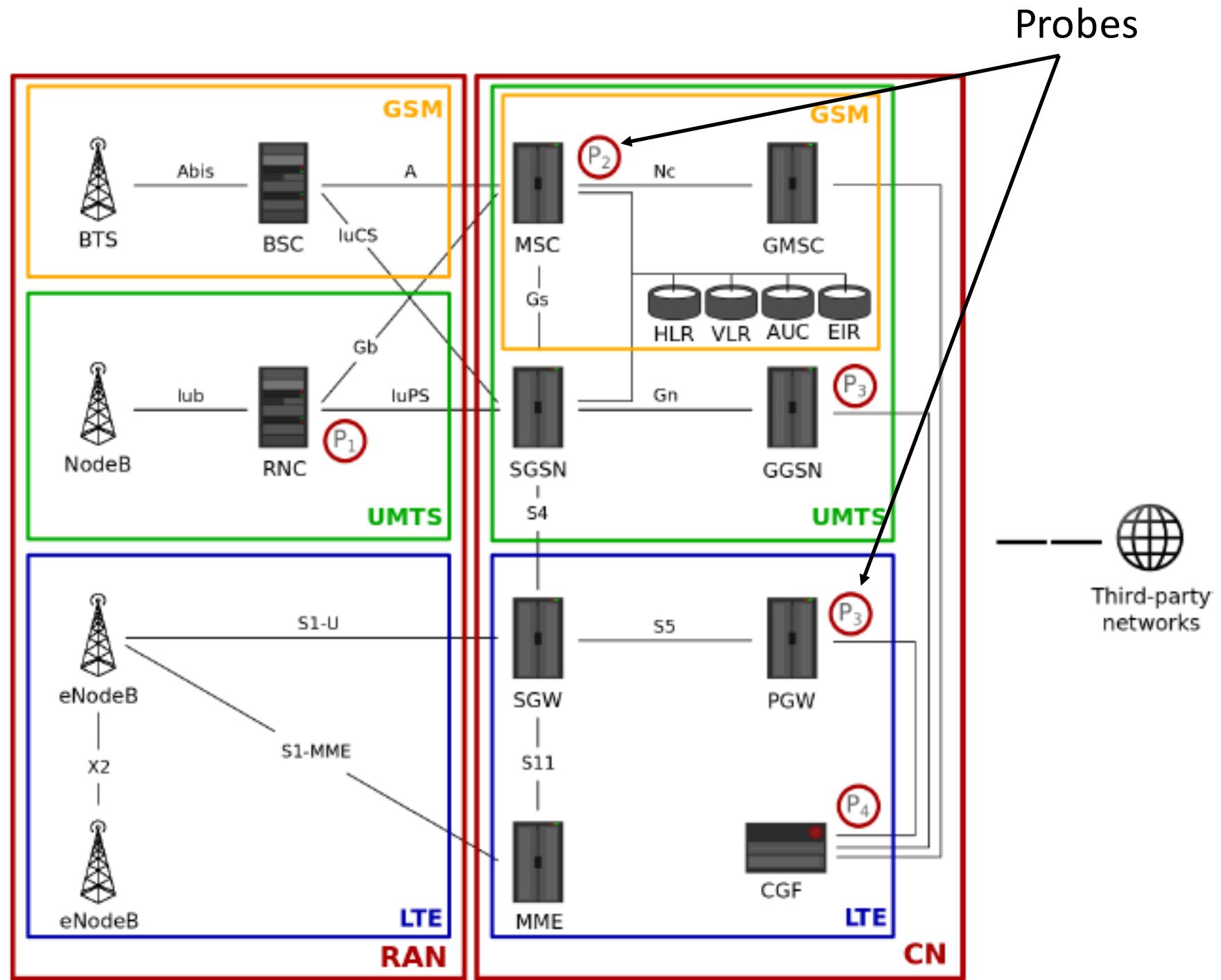
- Capacity, latency, energy, …

Need of understanding, characterizing the mobile Internet traffic

- Collection of mobile data

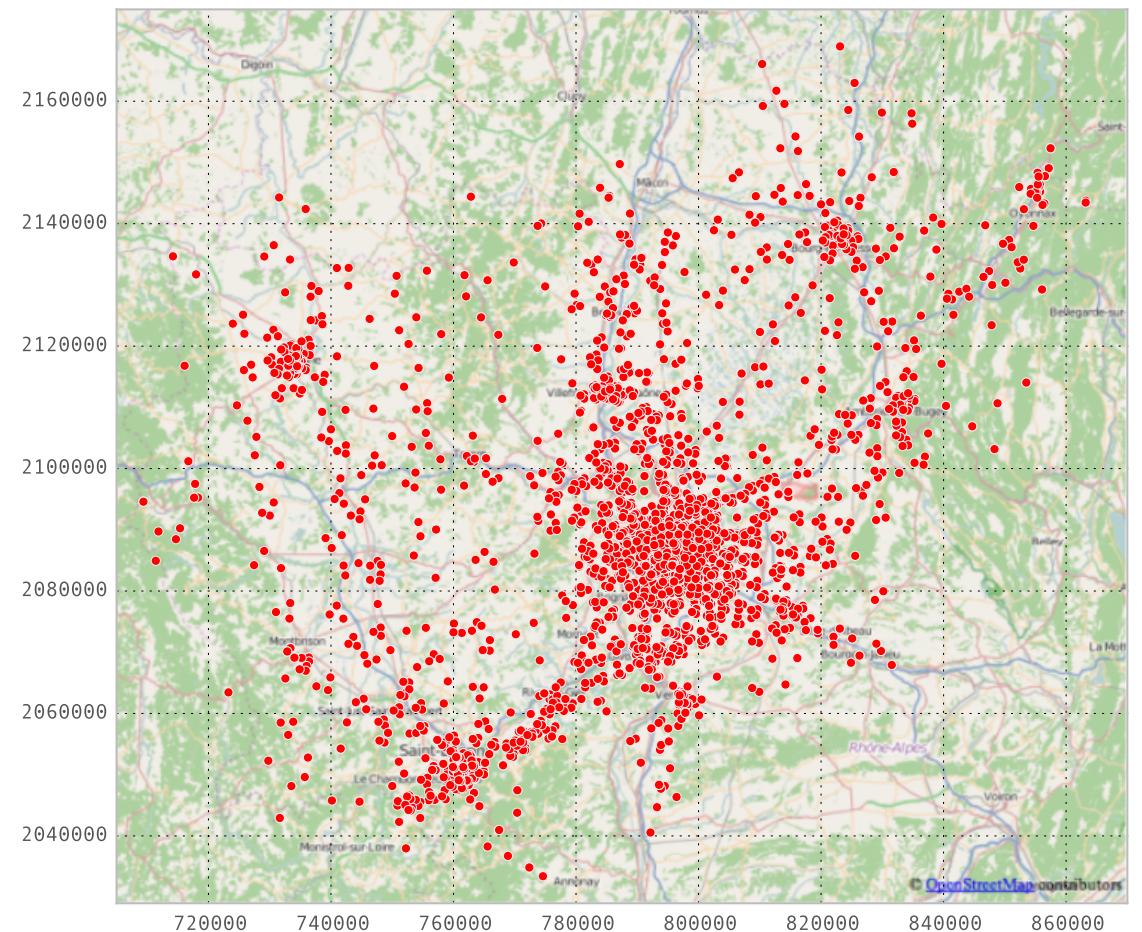- Use of data mining, statistic techniques

Context

- Collection of mobile data

Our dataset

- A regional fine-grained dataset

    - Covers a large region around Lyon

    - Covers lots of Internet services

        - Web, P2P, Download, News,
        Mail, DB, Control, Games,
        Streaming, Chat, VOIP, VPN,
        MMS, …

    - 2G/3G/4G

    - Duration 14 days

    - Time granularity 6 minutes

Our dataset

6722 cellules

| period | lac | ci | sectorx | sectory | x | y | accesstype | density | nByte | nByteTCP | nByteUDP | Byte_Web | Byte_Streaming | …
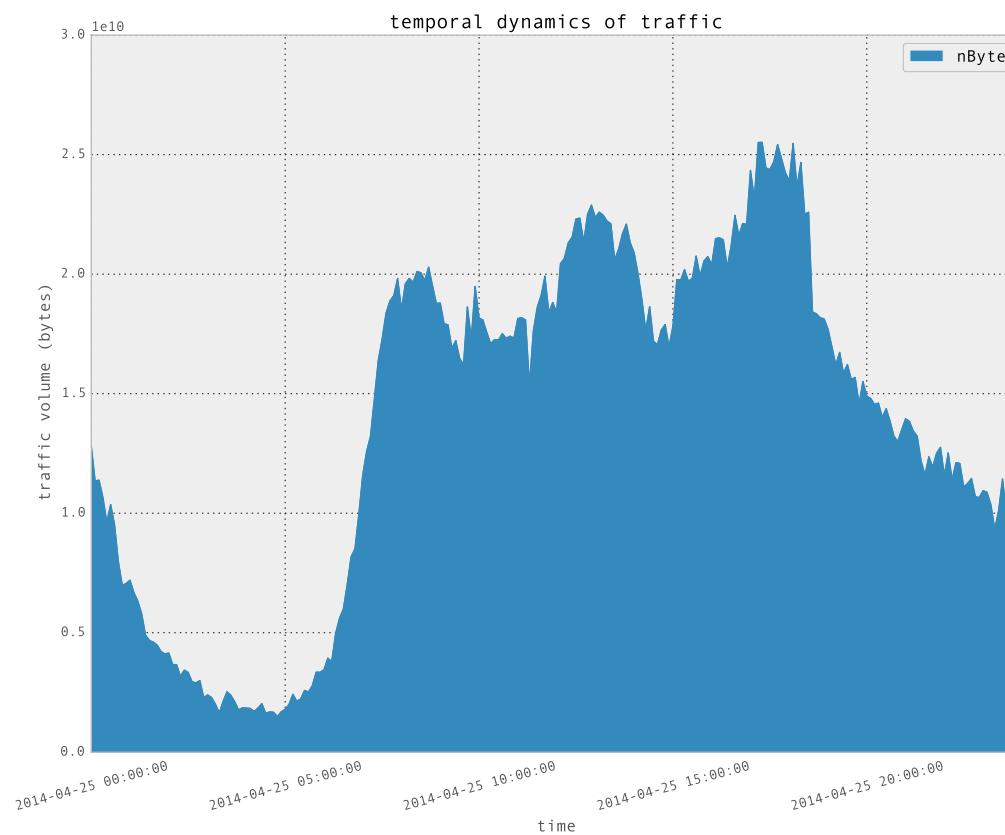
Positions of cellules
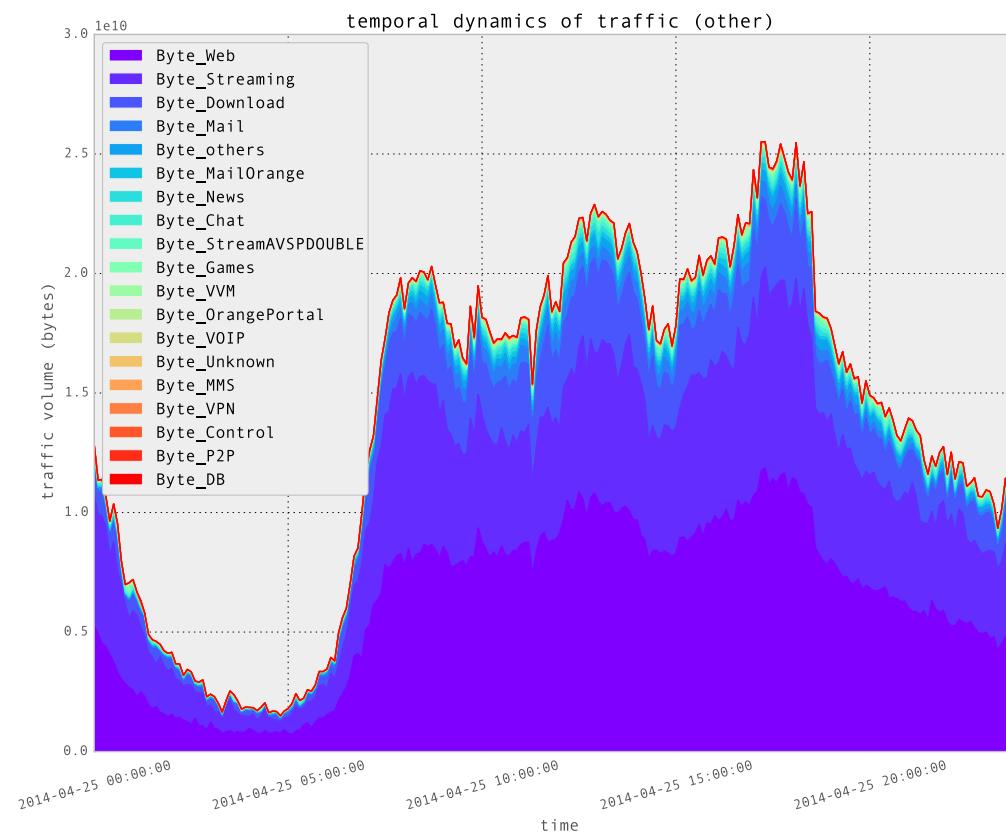
20 services

Number of users

Time (each 6 minutes)   2G, 3G, 4G
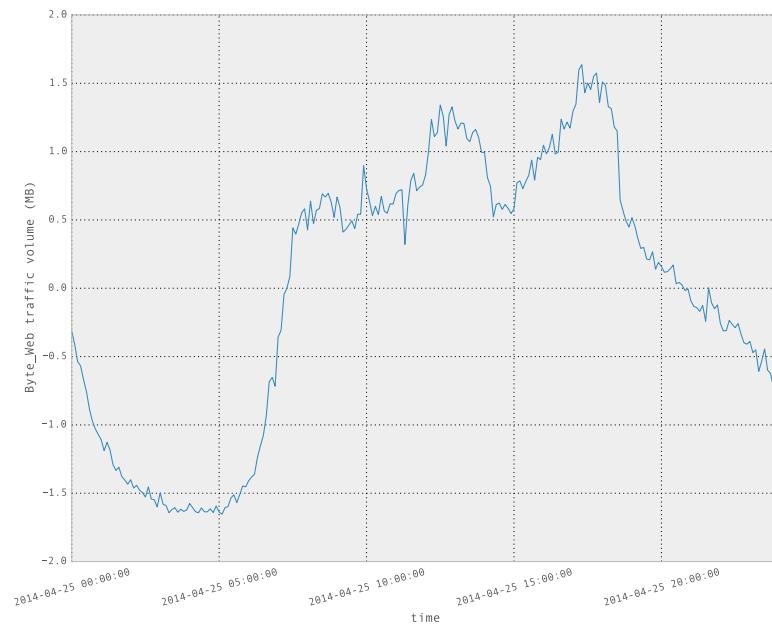
Visualisation & analysis

Temporal dynamics of traffics

Aggregate traffic

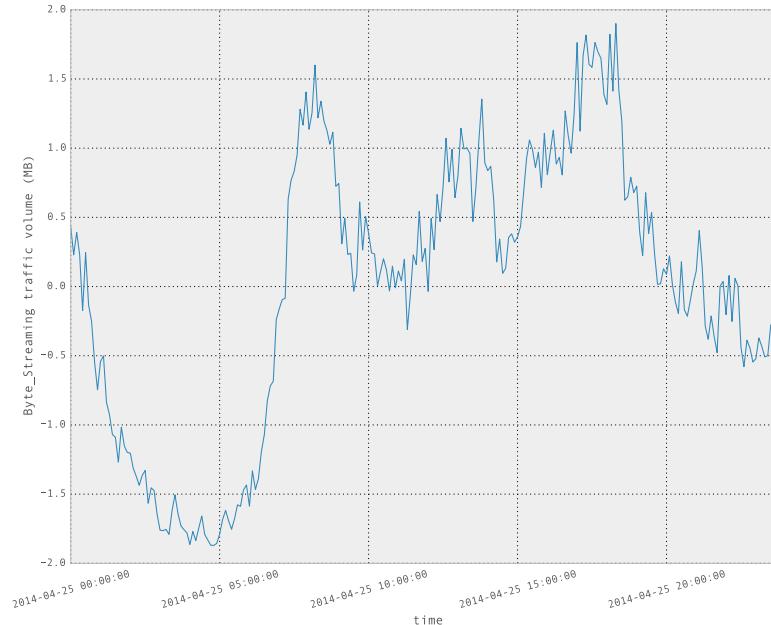Per-service traffics (stacked)

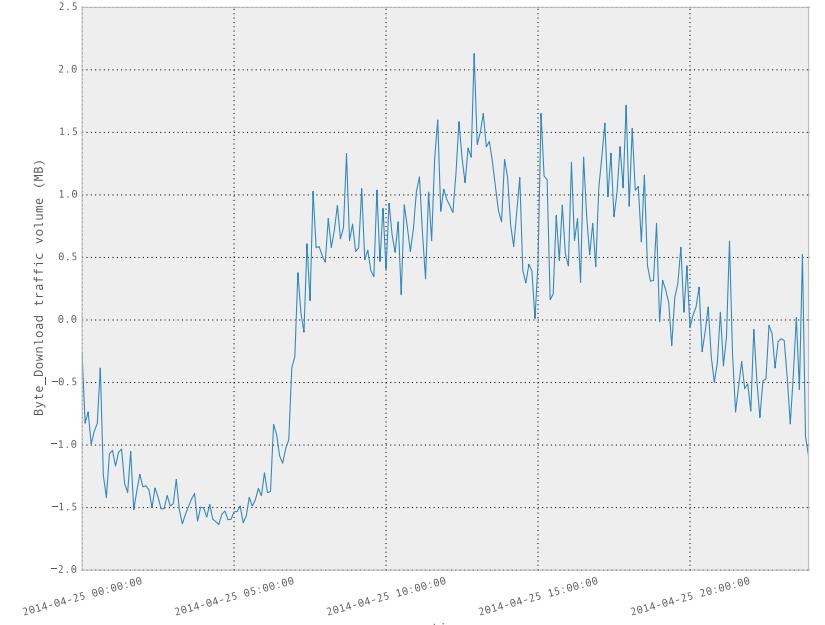Anh-Dung NGUYEN - UrbaNet

Visualisation & analysis
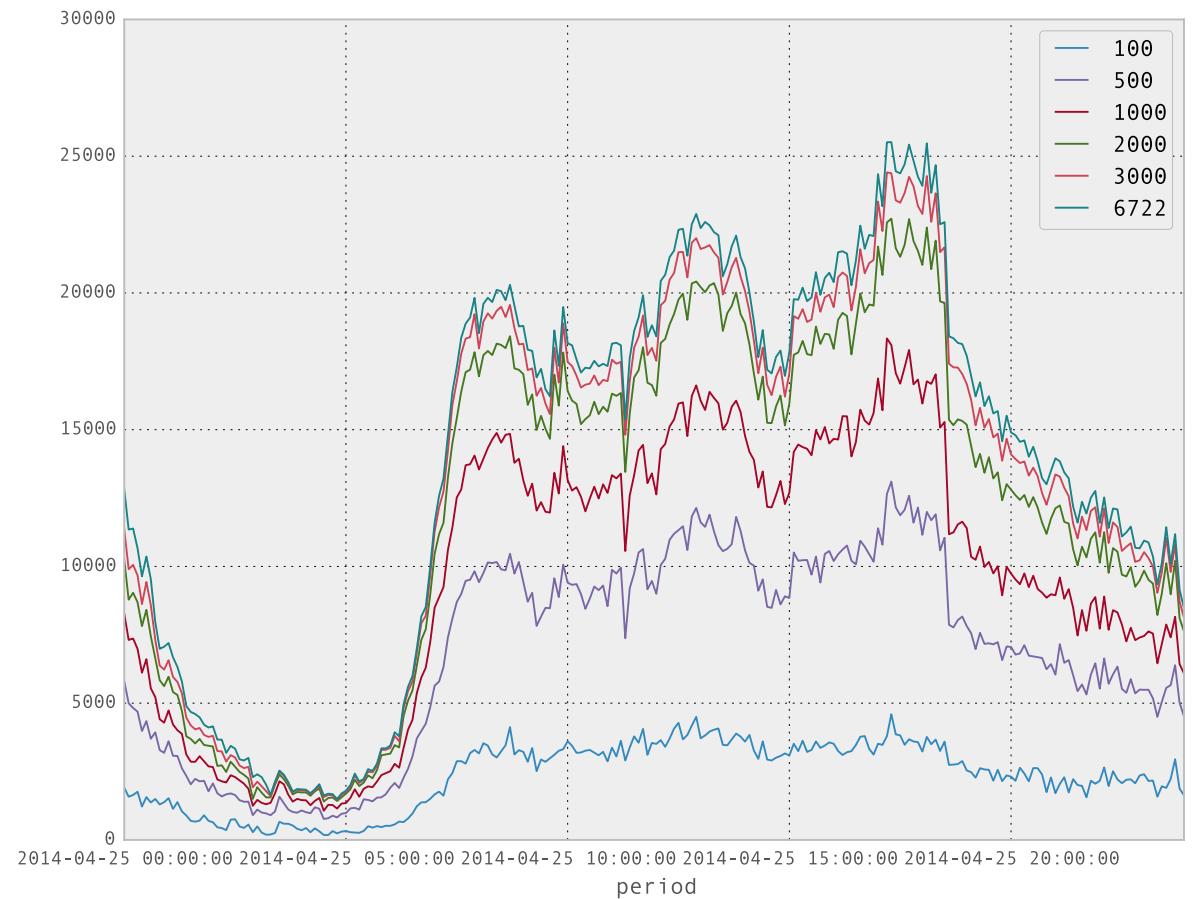
Temporal dynamics of traffics

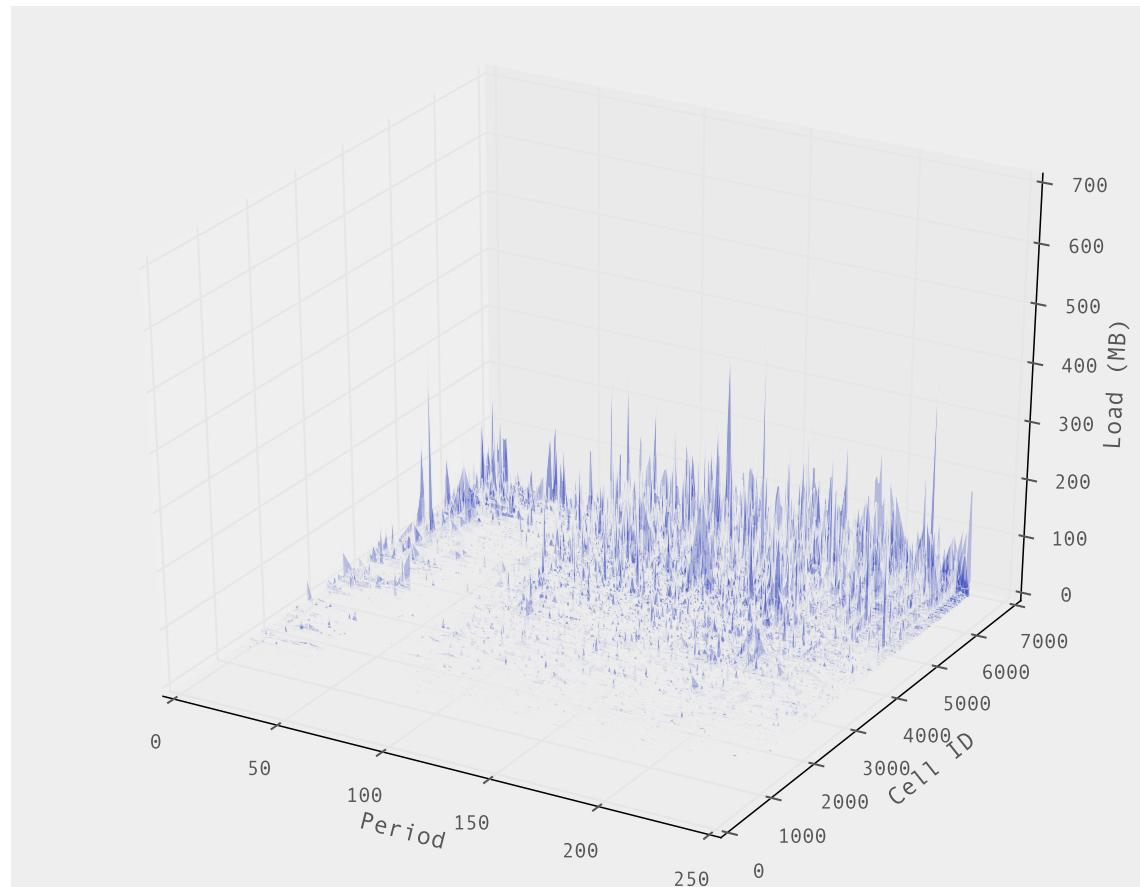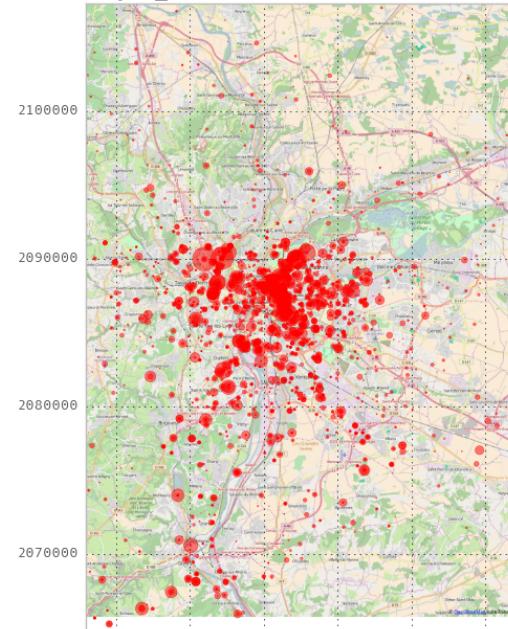Web   Streaming   Download

Visualisation & analysis
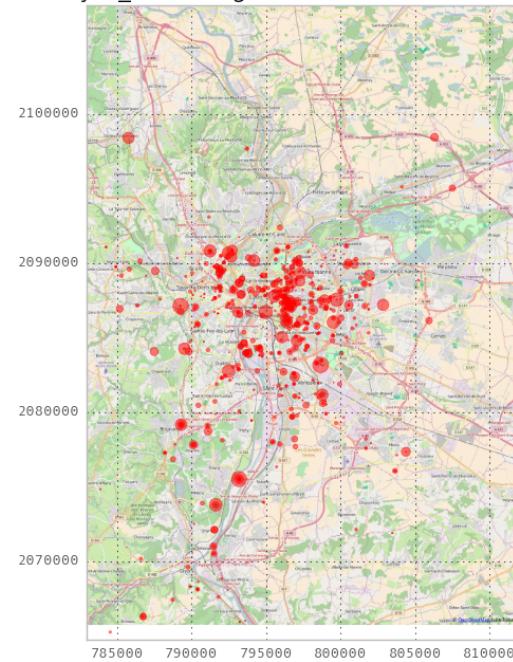
Traffic at each base station

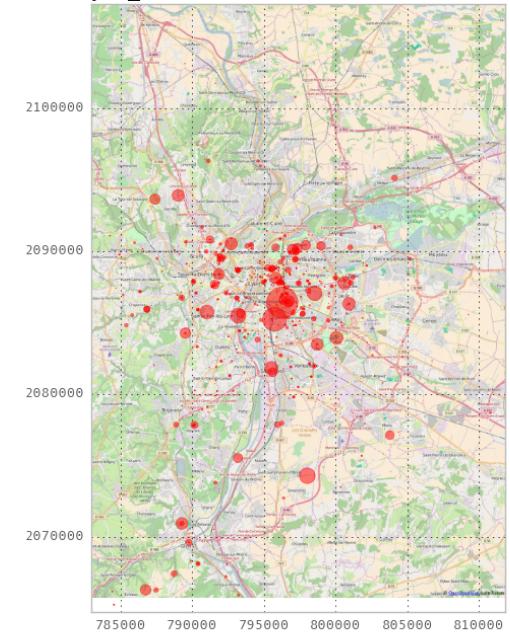# Visualization & analysis

Web

Byte_Web at 2014-04-25 12:00:00

Streaming

Byte_Streaming at 2014-04-25 12:00:00
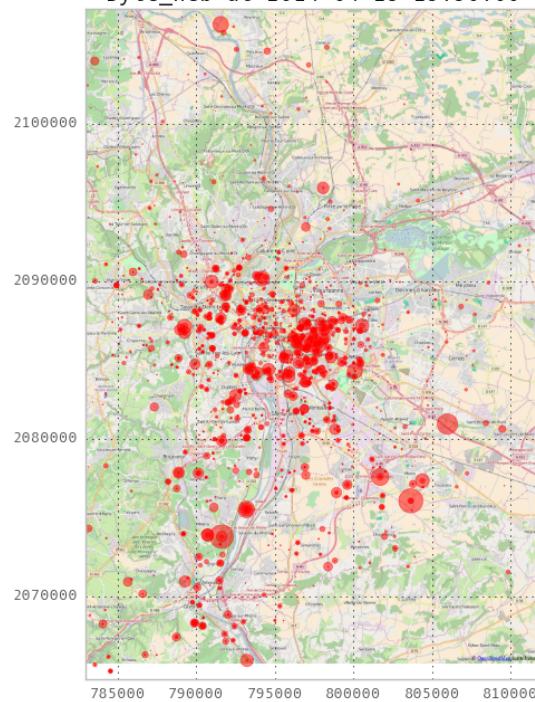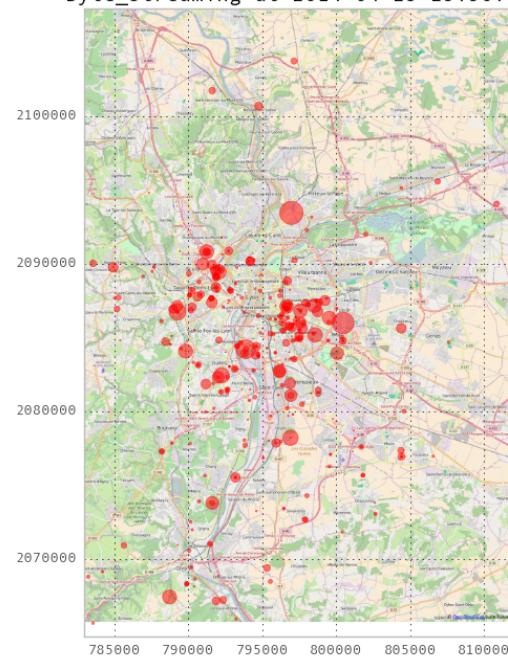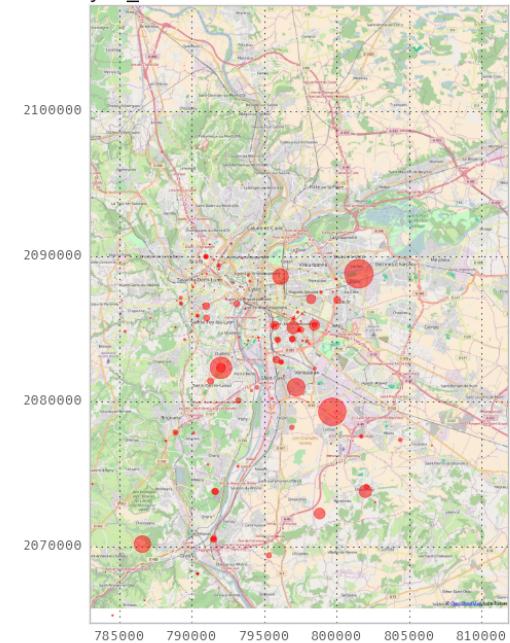
Download

Byte_Download at 2014-04-25 12:00:00

12h

Byte_Web at 2014-04-25 23:30:00

Byte_Streaming at 2014-04-25 23:30:00

Byte_Download at 2014-04-25 23:30:00
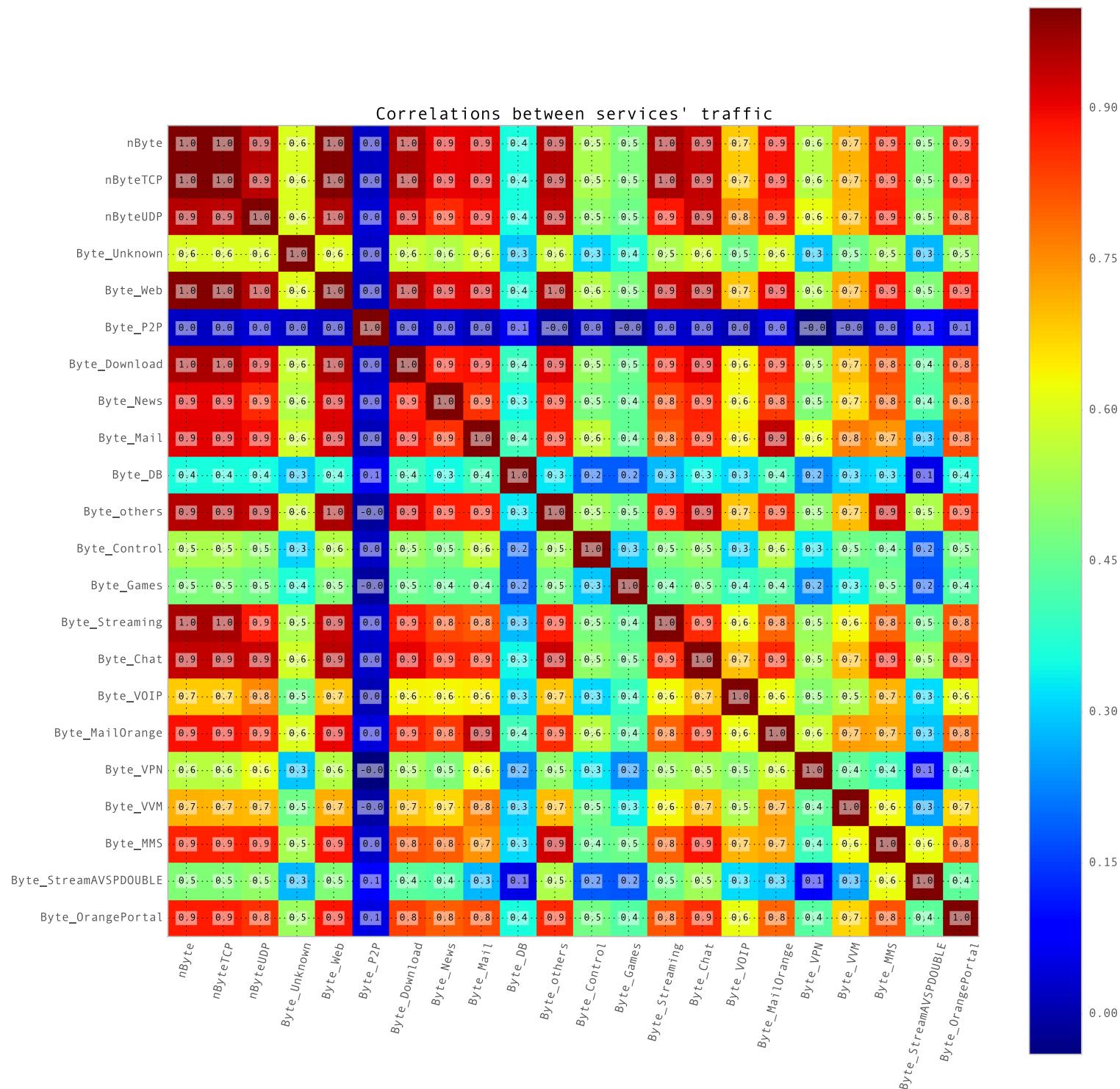
23h30

Correlation between aggregate traffics
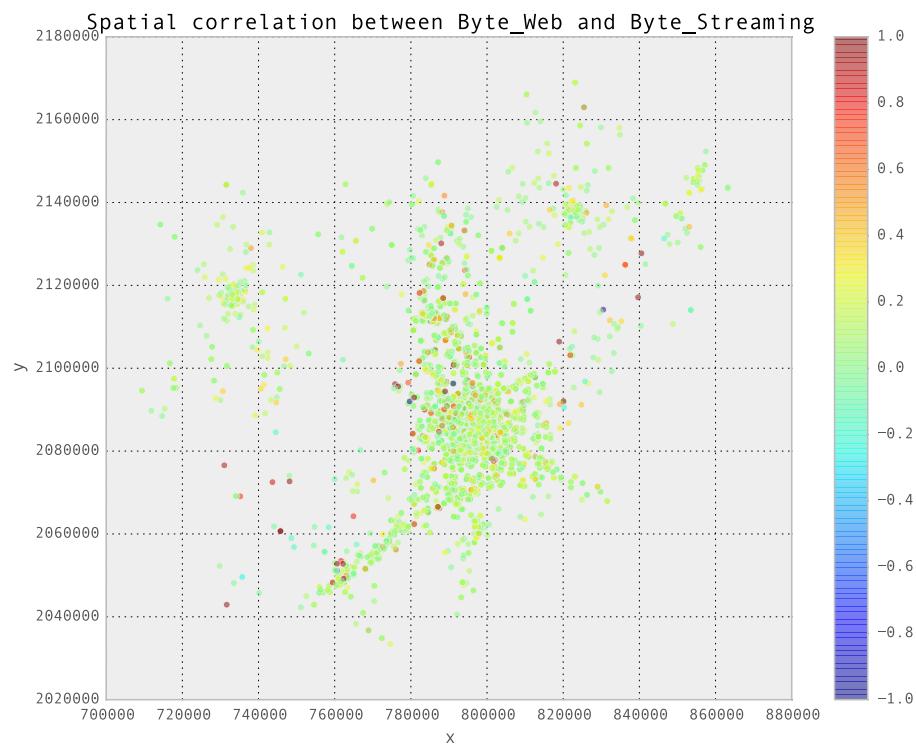

Correlations between services' traffic

Dominant traffics
are
highly correlated!

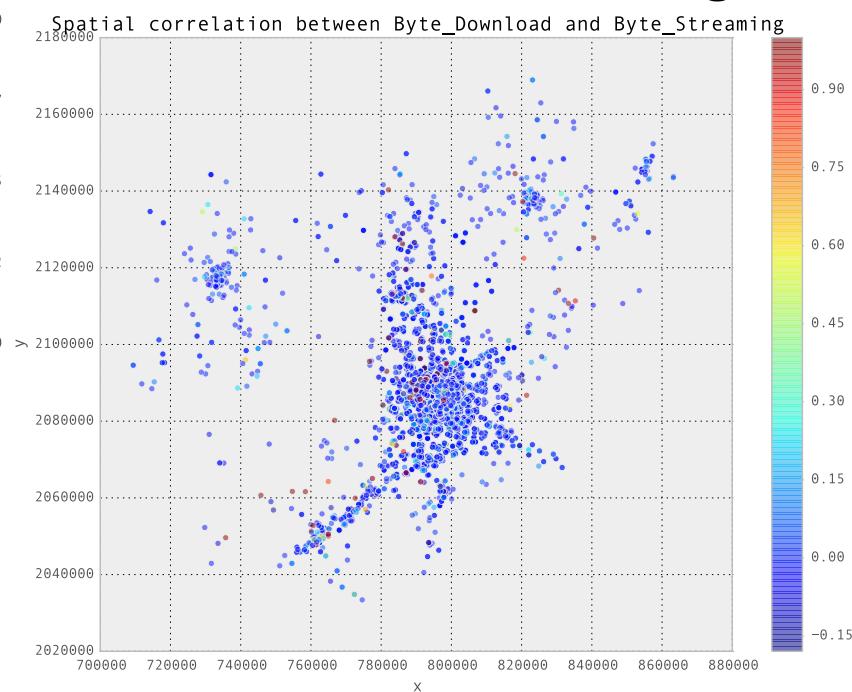Anh-Dung NGUYEN - UrbaNet

Correlation between individual traffics

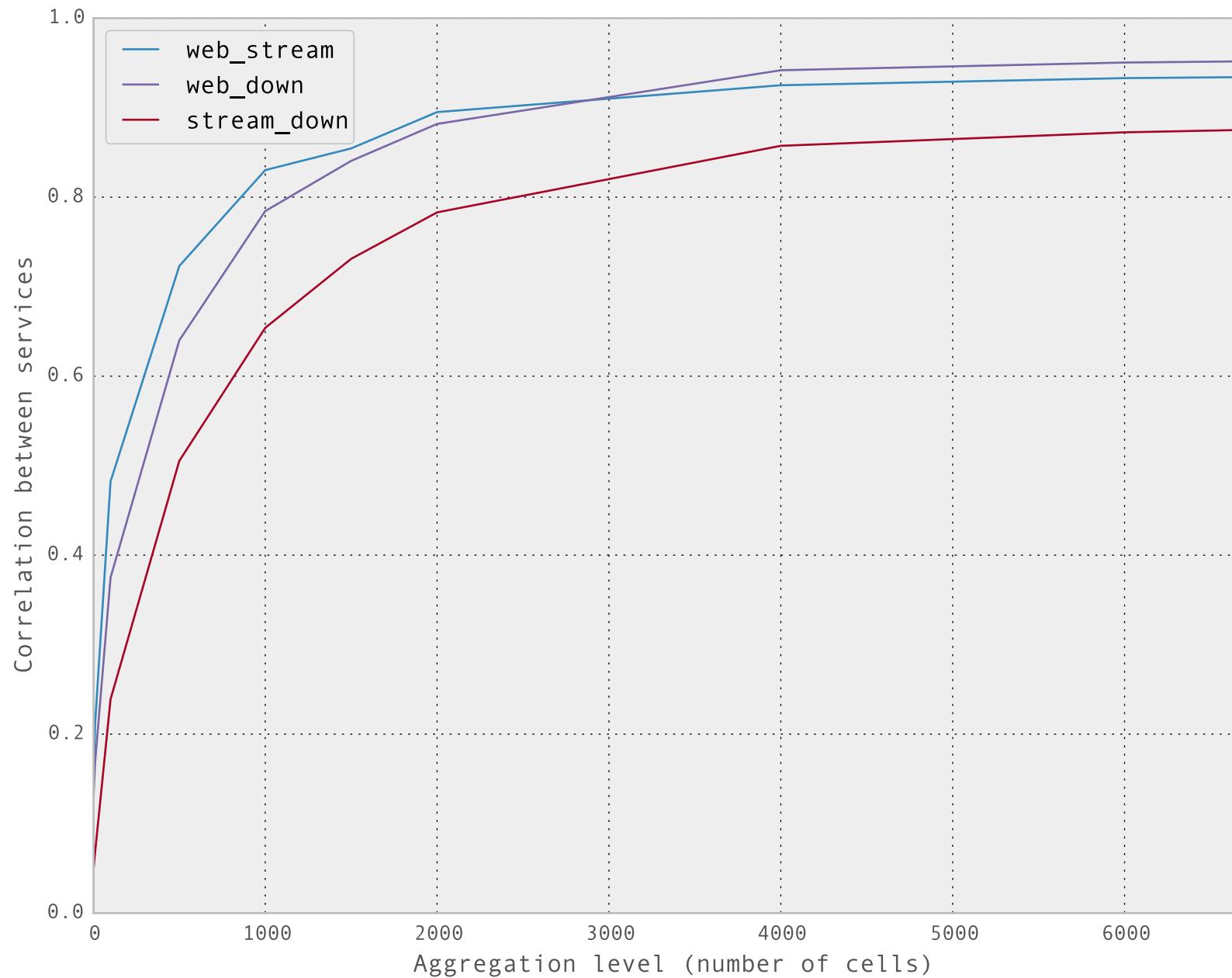
Web vs Streaming

Web vs Download
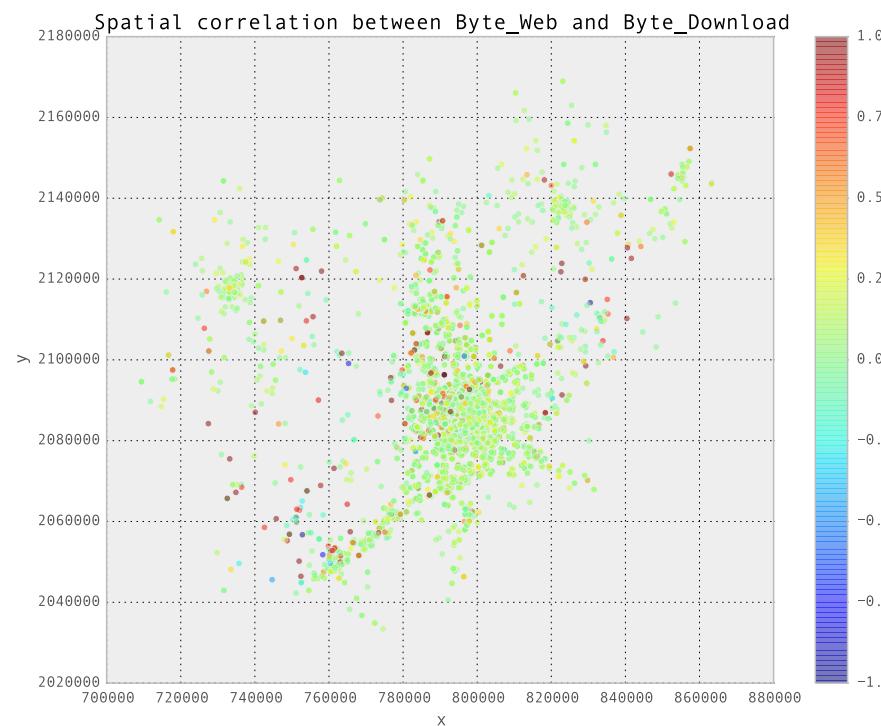
Download vs Streaming

1 base station = low correlation, all base stations = strong correlation

Simpson paradox?

Simpson paradox

Correlation as a function of aggregation level

Aggregation level = Number of cells whose traffics are aggregated

Conclusion & Perspectives

- Need to understanding the spatio-temporal dynamics of mobile Internet traffic

- A fine-grained dataset to study mobile traffic

- Non-trivial dynamics of traffics at aggregate/individual level

- Aggregate vs disaggregate view → Simpson paradox

- Push further the study on the spatio-temporal correlation between services

- Metrics for characterizing the spatio-temporal dynamics of traffics

- Study of the correlation between mobile Internet usages and mobility of users

Thank you!

NOM DU CHAPITRE