Leveraging CDR datasets for Context-Rich Performance Modeling of Large-Scale Mobile Pub/Sub Systems

Georgios Bouloukakis, Rachit Agarwal, Nikolaos Georgantas, Animesh Pathak, Valérie Issarny

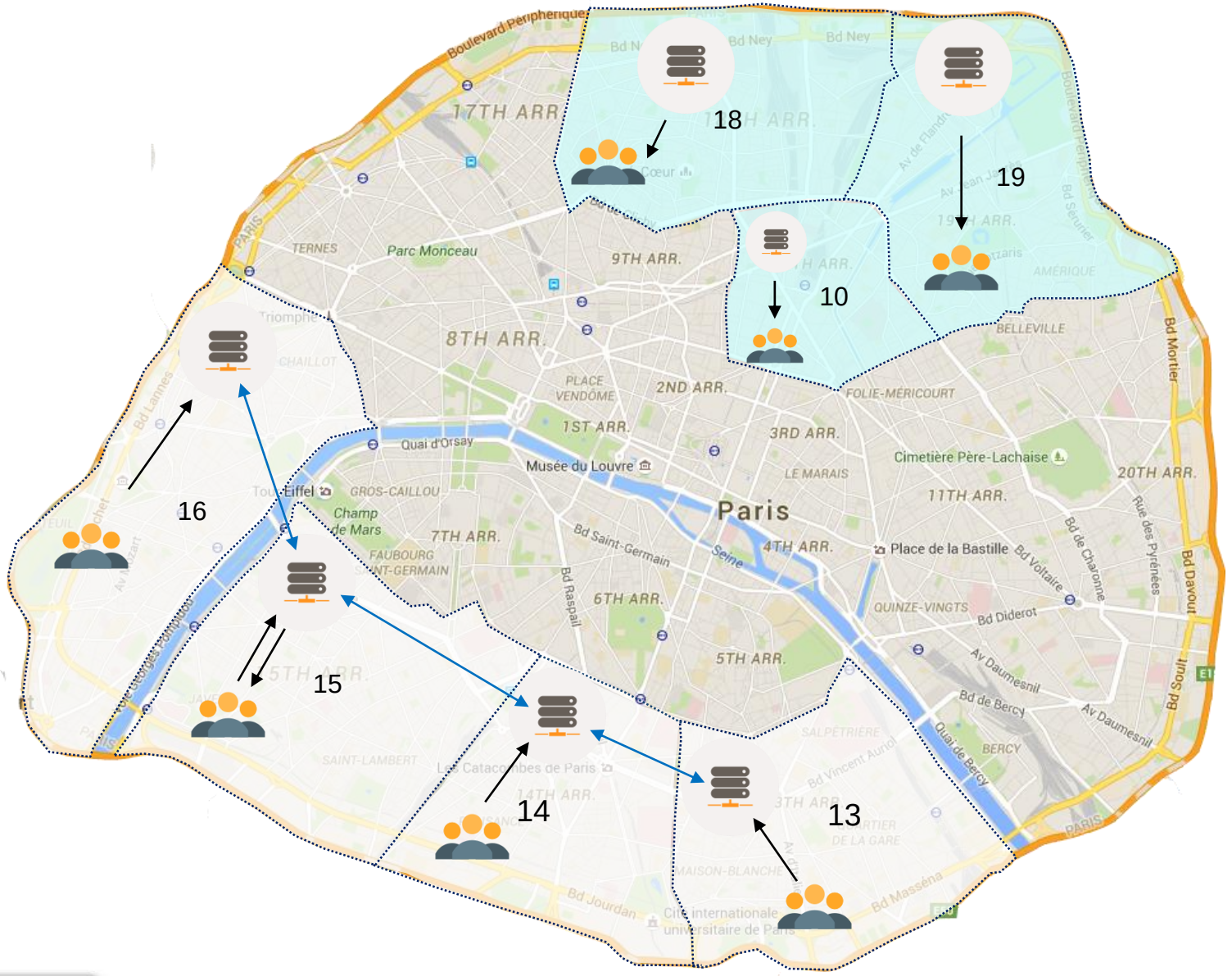
WiMob, Abu Dhabi, October 2015

MiMove project-team
Inria Paris-Rocquencourt, France

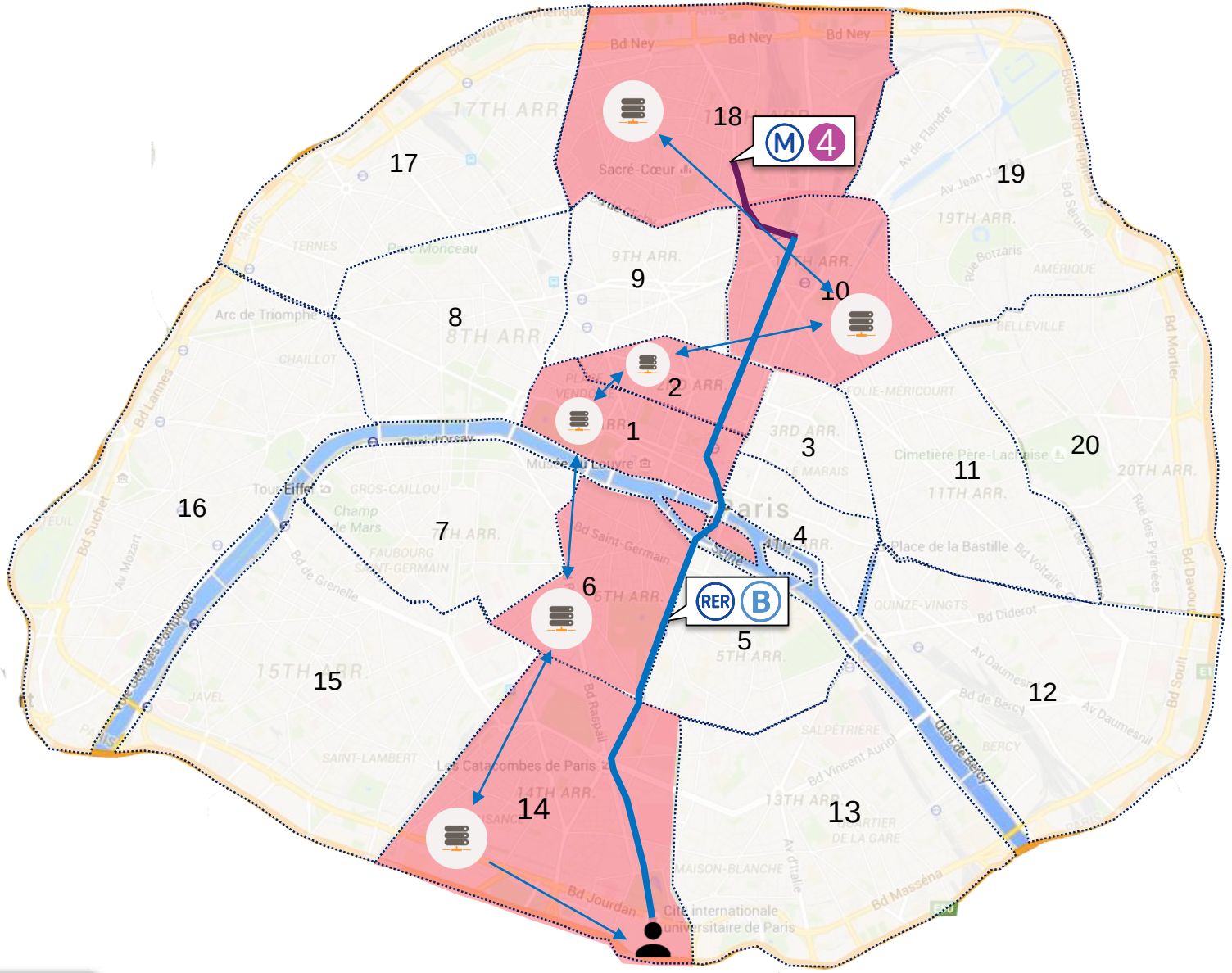
MiMove
Middleware on the Move

Inria

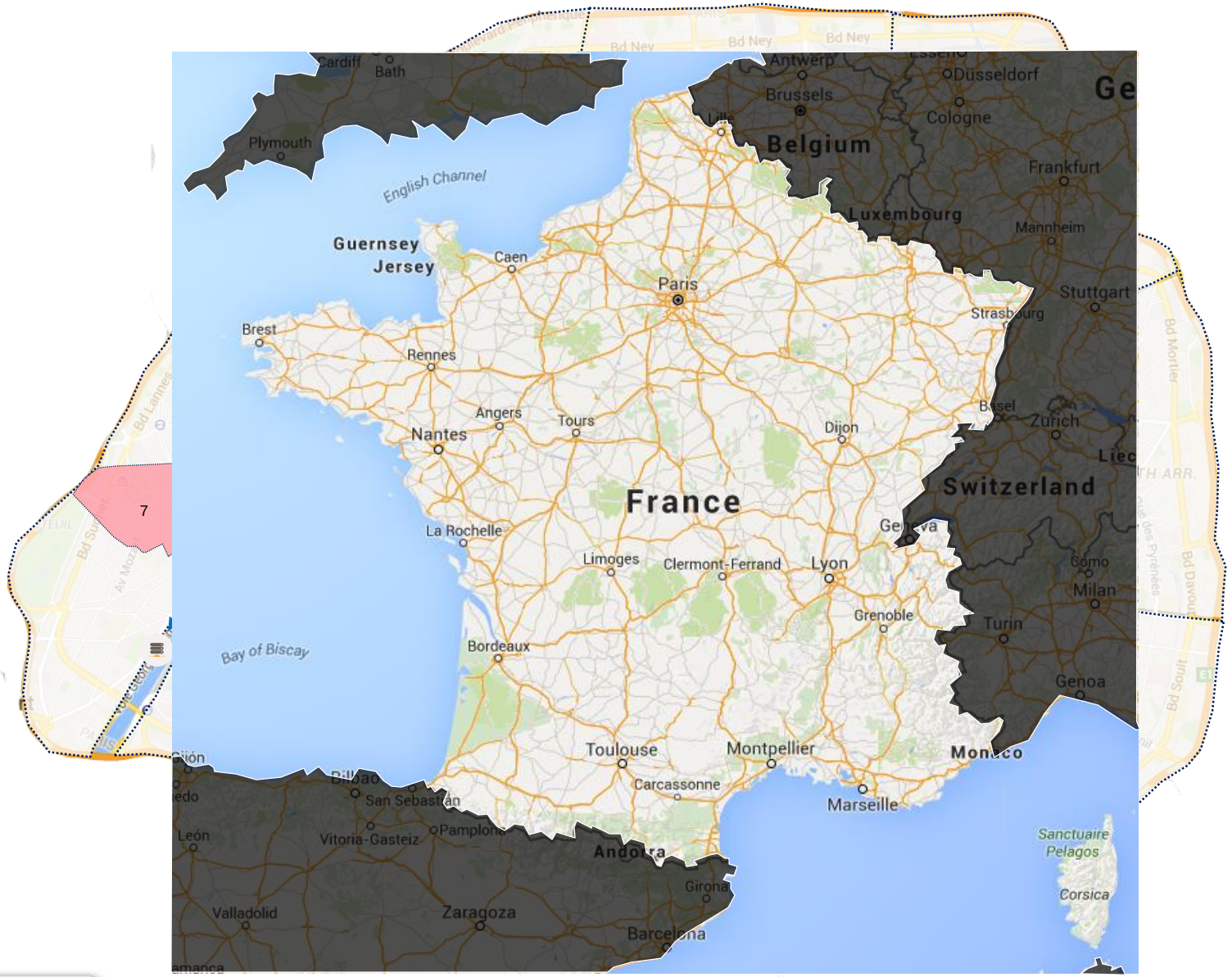
Motivation


18
19
10
16
15
14
13
Paris
17TH ARR.
9TH ARR.
8TH ARR.
2ND ARR.
1ST ARR.
3RD ARR.
7TH ARR.
6TH ARR.
5TH ARR.
4TH ARR.
11TH ARR.
20TH ARR.
Parc Monceau
Musée du Louvre
Champ de Mars
Quai d'Orsay
Cimetière Père-Lachaise
Place de la Bastille
Seine
Bd Saint-Germain
Bd Raspail
Bd Diderot
Av Daumesnil
Bd de Bercy
Quai de Bercy
Bd Davout
Bd Masséna
Bd Jourdan
Bd Mortier
Bd Voltaire
Bd de Charonne
Rue des Pyrénées
Boulevard Périphérique
TERNES
PLACE VENDÔME
FOLIE-MÉRICOURT
BELLEVILLE
LE MARAIS
QUINZE-VINGTS
BERCY
SAINT-LAMBERT
SALPÊTRIÈRE
GROS-CAILLOU
FAUBOURG ST-GERMAIN
CHAILLOT
AMÉRIQUE
Cité internationale universitaire de Paris

Motivation



Motivation

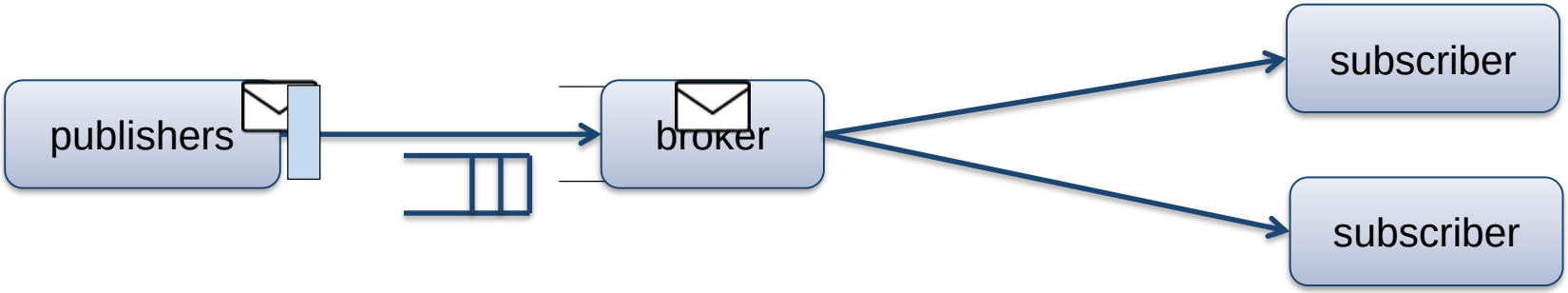

Cardiff
Bath
Plymouth
English Channel
Guernsey
Jersey
Caen
Brest
Rennes
Angers
Tours
Nantes
La Rochelle
Bay of Biscay
Bordeaux
Limoges
Clermont-Ferrand
Lyon
Grenoble
Dijon
Paris
France
Antwerp
Brussels
Belgium
Lille
Düsseldorf
Cologne
Frankfurt
Luxembourg
Mannheim
Stuttgart
Strasbourg
Basel
Zurich
Switzerland
Geneva
Como
Milan
Turin
Genoa
Monaco
Marseille
Montpellier
Toulouse
Carcassonne
Andorra
Girona
Barcelona
San Sebastián
Vitoria-Gasteiz
Pamplona
León
Valladolid
Zaragoza
Sanctuaire Pelagos
Corsica
7

Outline

- **Performance Modeling**
- Large-scale Mobile Publish/Subscribe System
- CDRs Dataset Analysis for Senegal
- Context-rich Model Parameterization
- Simulation Results
- Conclusions

Performance Modeling

- We rely on Queueing Network Models (QNMs)
  - systems resources and networks are represented as *queues*
  - exchanged data are represented as *jobs* served at the queues
- Common ways to evaluate the performance:
  1. using existing closed-form solutions and probability distributions
  2. *performing simulations by analyzing more complex and realistic systems*
- Modeling a pub/sub system:



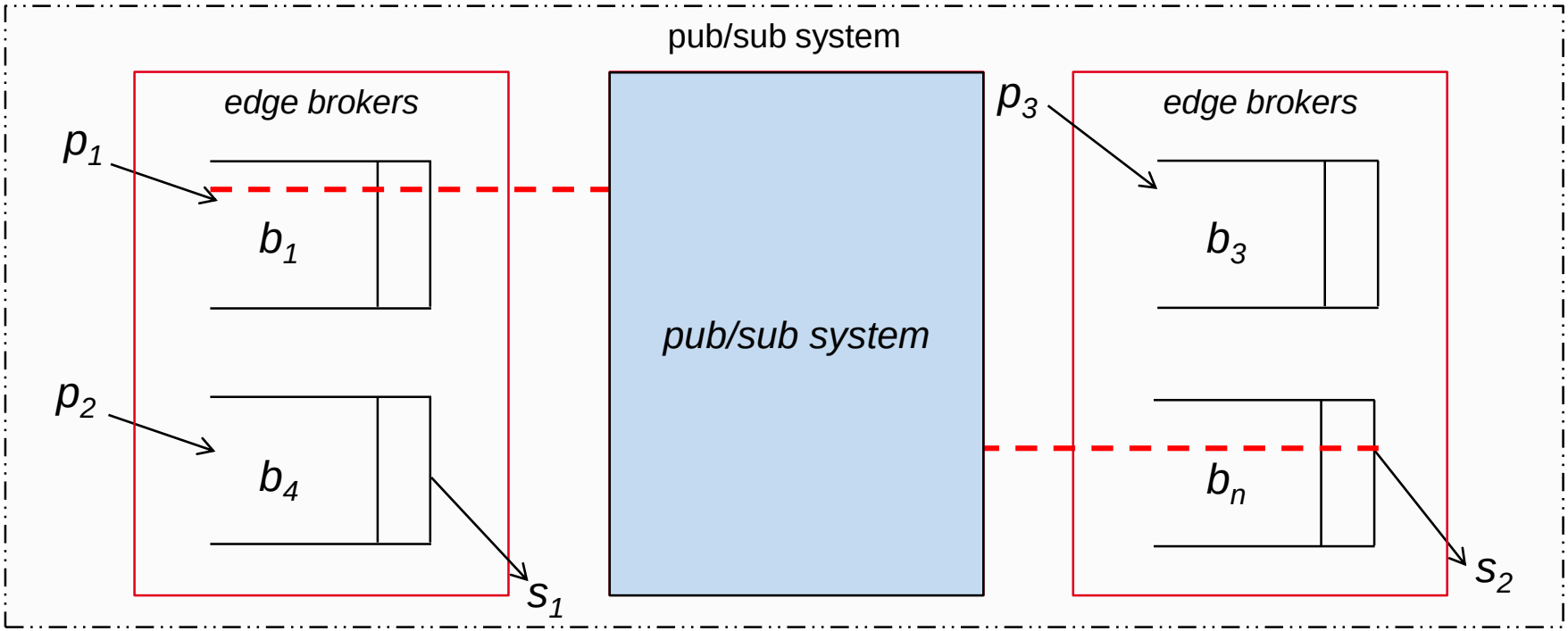
Outline

- Performance Modeling
- **Large-scale Mobile Publish/Subscribe System**
- CDRs Dataset Analysis
- Context-rich Model Parameterization
- Simulation Results
- Conclusions

Large-scale Mobile Pub/Sub System (1)

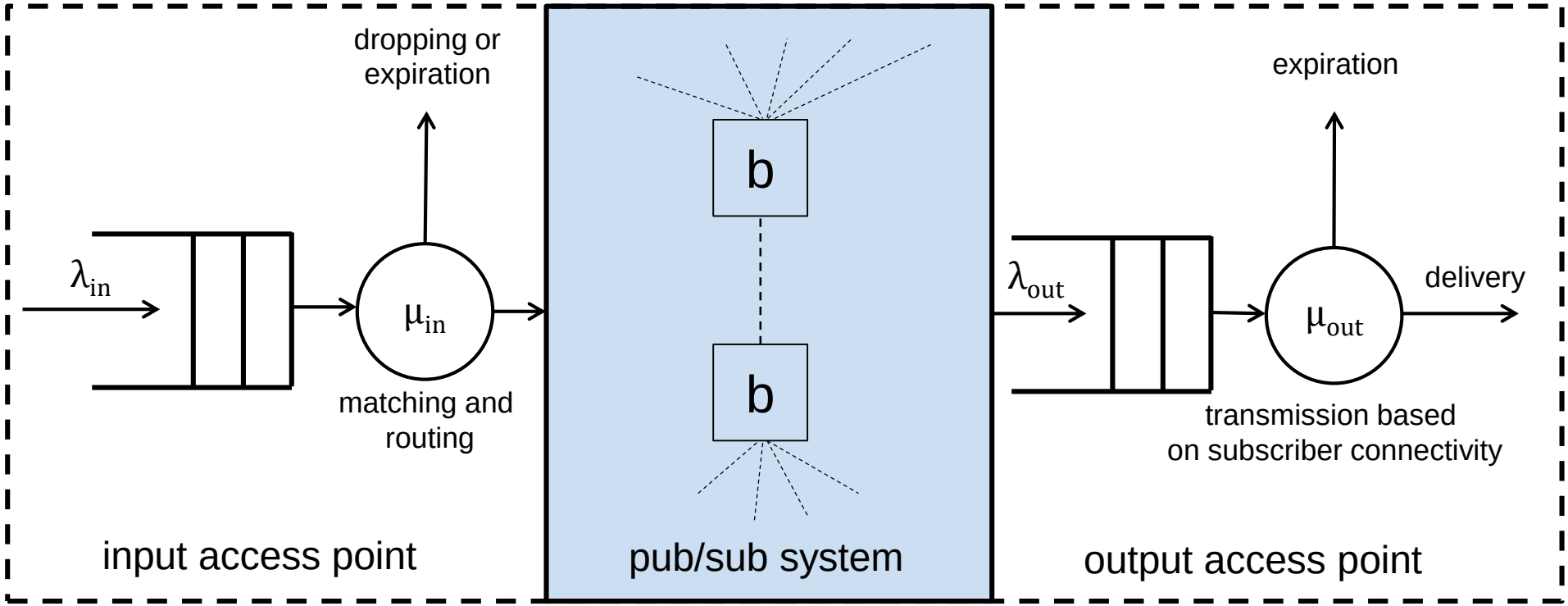
- Publishers (p):
  - mobile entities
  - they connect to publish events
  - a *lifetime* limit can be assigned to each event
- Subscribers (s):
  - mobile entities
  - they connect to system occasionally to receive events
  - they disconnect to save energy

- End-to-end interaction between p_1 and s_2:



Large-scale Mobile Pub/Sub System (2)

- End-to-end interaction model:



end-to-end response times are higher due to users' intermittent connectivity!

Outline

- Performance Modeling
- Large-scale Mobile Publish/Subscribe System
- **CDRs Dataset Analysis**
- Context-rich Model Parameterization
- Simulation Results
- Conclusions

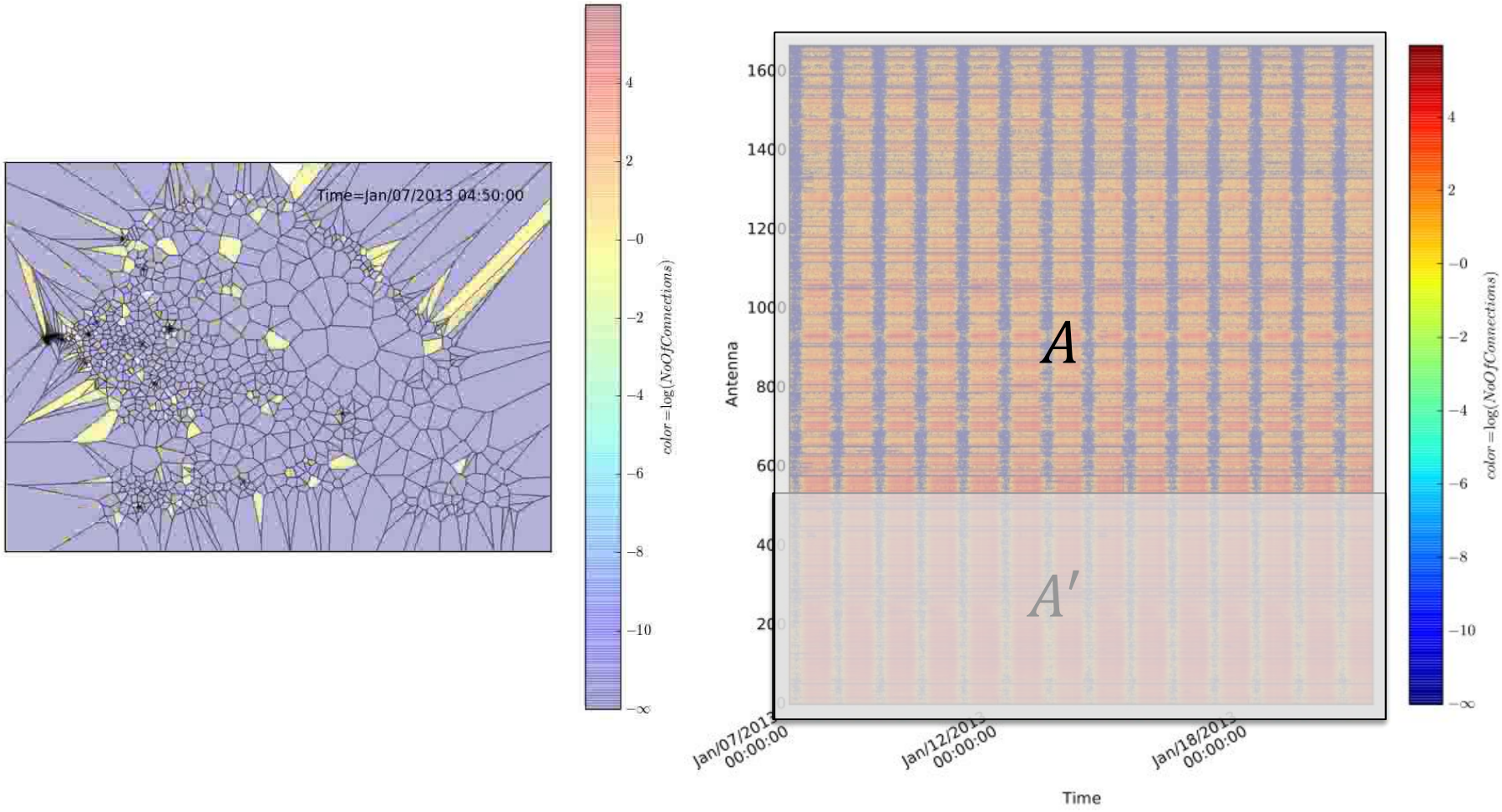
D4D Dataset

- D4D Dataset:
  - Generated by Orange labs for the subscribers of Sonatel Network in **Senegal**
  - Contains Call Detail Records (CDRs)
  - Collected over 50 weeks starting from 7th January 2013
  - Every 10 min interval, the location of the associated antenna is recorded when a user makes a call or sends an sms

- CDRs for parameterizing our model:
  - user access to mobile services is similar with user access to application services
  - they reflect location and time context across the whole country

D4D Dataset Analysis

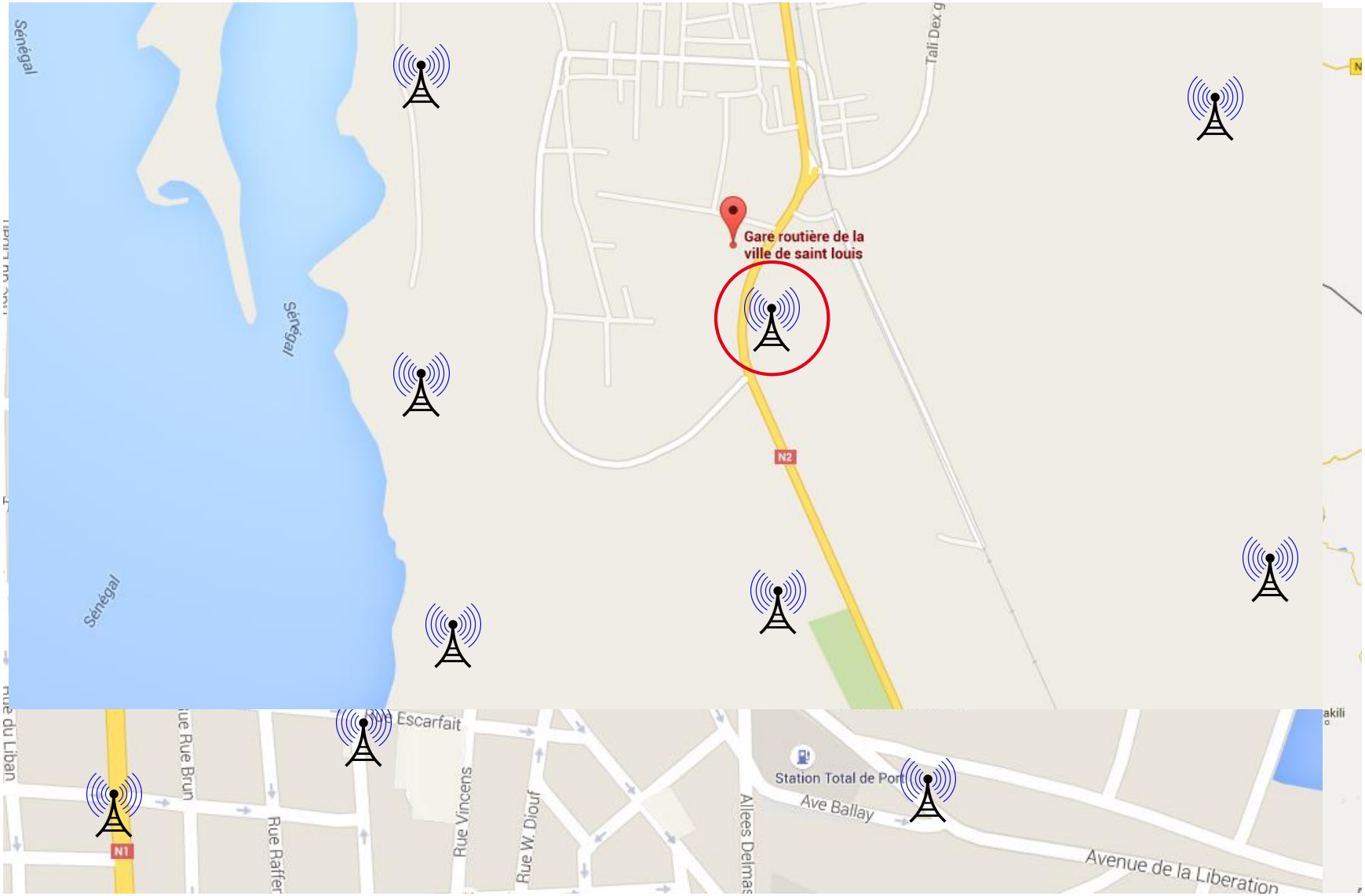
Orange Data for Development Challenge in Senegal



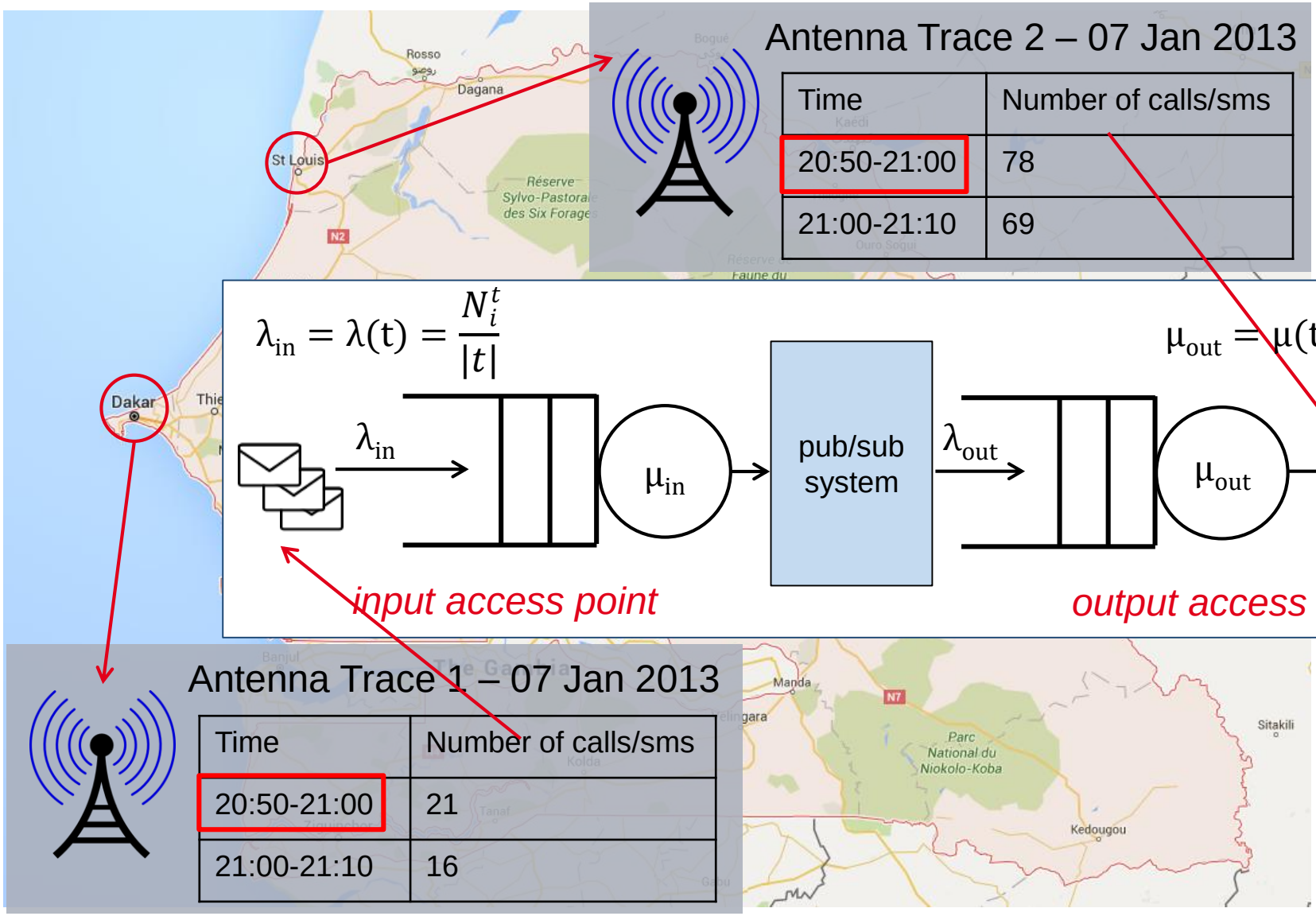
Outline

- Performance Modeling
- Large-scale Mobile Publish/Subscribe System
- CDRs Dataset Analysis
- **Context-rich Model Parameterization**
- Simulation Results
- Conclusions

Context-rich Model Parameterization



Context-rich Model Parameterization



Antenna Trace 2 – 07 Jan 2013

Time	Number of calls/sms
20:50-21:00	78
21:00-21:10	69

$$\lambda_{in} = \lambda(t) = \frac{N_i^t}{|t|}$$

$$\mu_{out} = \mu(t) = \frac{N_j^t}{|t|}$$

input access point

output access point

Antenna Trace 1 – 07 Jan 2013

Time	Number of calls/sms
20:50-21:00	21
21:00-21:10	16

Outline

- Performance Modeling
- Large-scale Mobile Publish/Subscribe System
- CDRs Dataset Analysis
- Context-rich Model Parameterization
- **Simulation Results**
- Conclusions

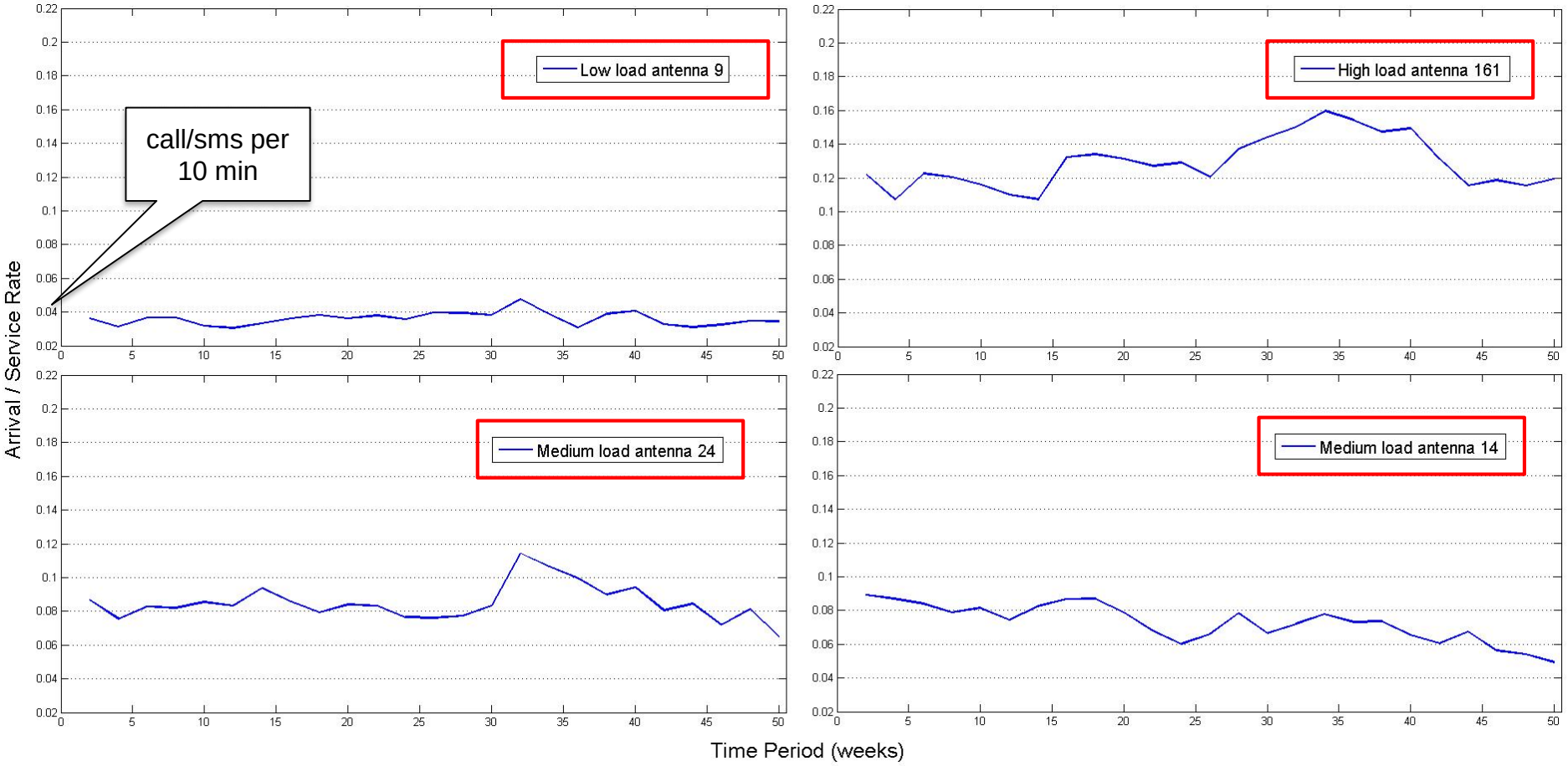
Simulation Results

- MobileJINQS[1]:
  - open source simulator for Queueing Network Models
- Using MobileJINQS we:
  - implement our pub/sub model
  - evaluate the response time by applying:
    - Incoming loads and service delays of realistic traces from the D4D dataset
    - appropriate lifetime periods
- We classify the load of varied antenna traces into three categories:
  1. *low load antenna*
  2. *medium load antenna*
  3. *high load antenna*

[1] http://xsb.inria.fr/d4d#mobilejinqs

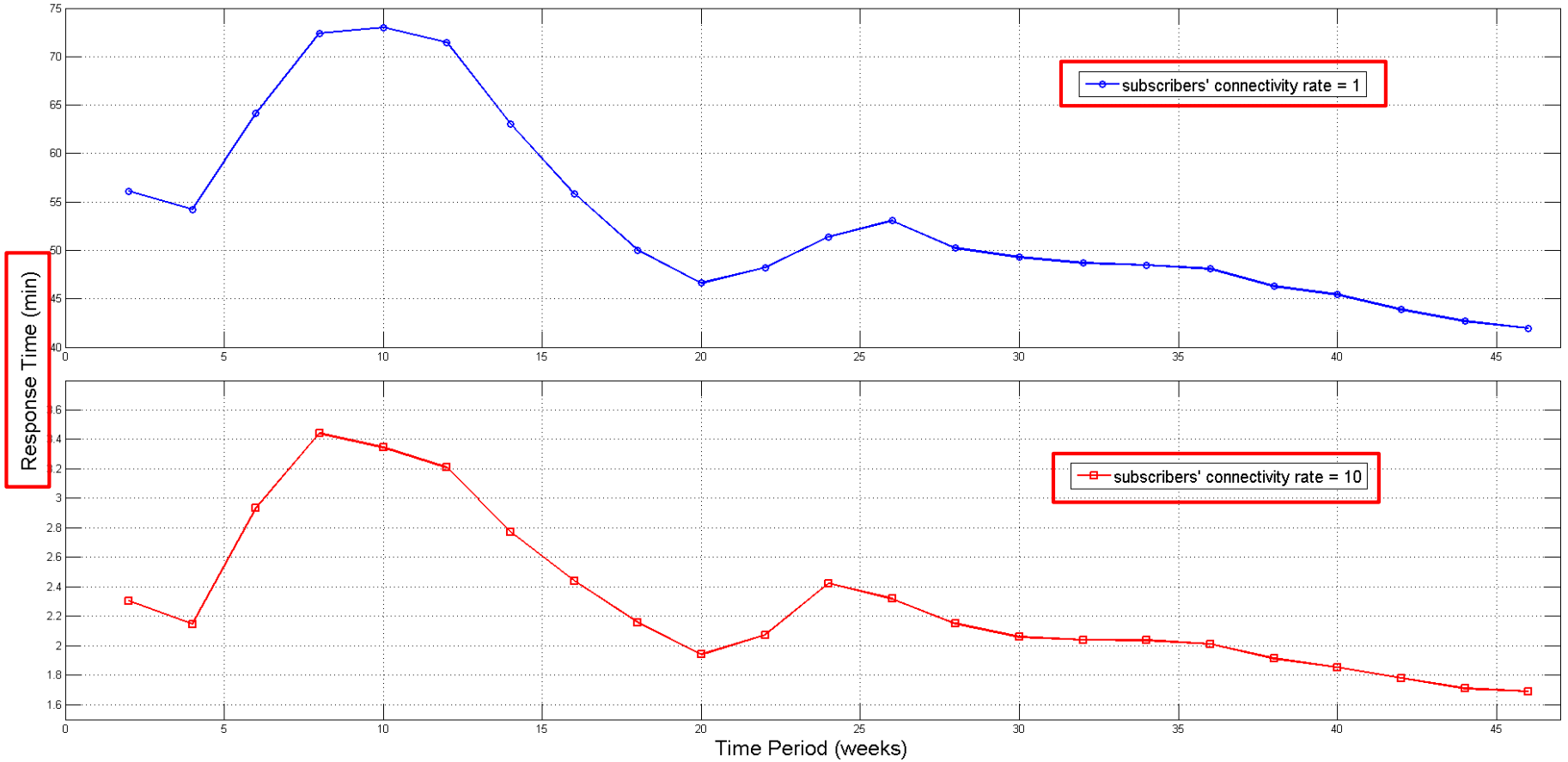
Simulation Results: representative input load

- Low, Medium and High load of antennas used for our experiments



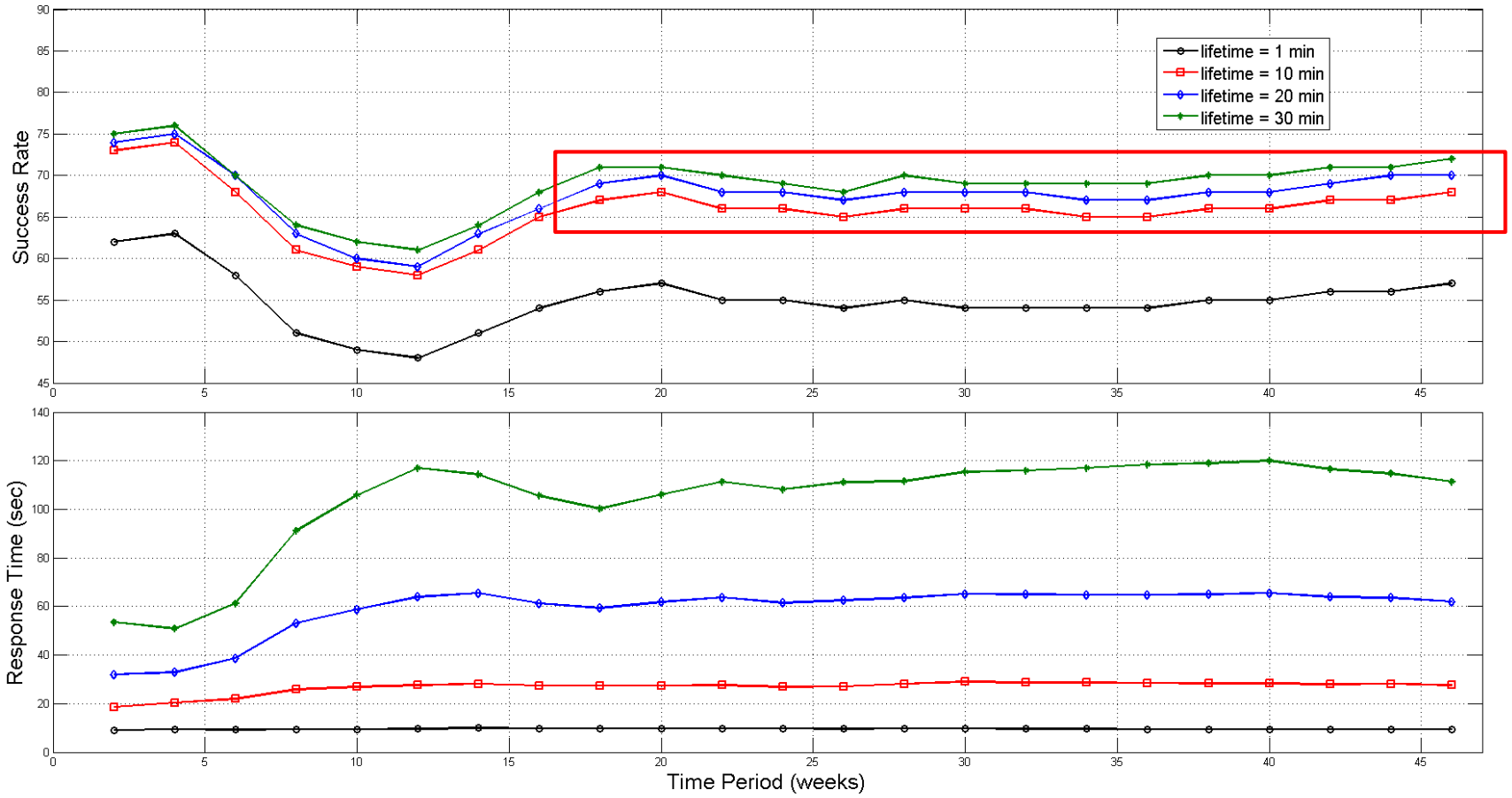
Simulation Results: response time for infinite lifetime

- End-to-end transactions from low load Antenna 9 to high load Antenna 161



Simulation Results: response time vs. delivery success rate with varying lifetime periods

- End-to-end transactions from low load Antenna 9 to high load Antenna 161



Outline

- Performance Modeling
- Large-scale Mobile Publish/Subscribe System
- CDRs Dataset Analysis
- Context-rich Model Parameterization
- Simulation Results
- **Conclusions**

Conclusion and future perspective

- We study the behavior of the underlying communication infrastructure for a mobile pub/sub system
- We evaluate the performance under context-rich realistic workload by utilizing CDRs from the D4D dataset
- By introducing varied lifetime periods we evaluate the trade-off between the delivery success rates and response times
- System designers should apply data validity lifetime periods depending on the context to achieve high performance

➢ Next step

- we are working on an analytical model for the estimation of response times. We intend to use the D4D dataset for the validation of it

Thank you

Further information:

Inria MiMove: mimove.inria.fr

D4D Challenge: xsb.inria.fr/d4d