AllAboard: a system for exploring mobility and optimizing transport in developing countries using cellphone data

M. Berlingerio, F. Calabrese, G. Di Lorenzo, R. Nair, F. Pinelli, M.L. Sbodio
IBM Research - Ireland

Public Information

SmarterCities Technology Centre

© 2013 IBM Corporation

Urbanization in **Growth Markets** and Challenges

Infrastructure and services have not kept pace with rapid rate of urbanization
Vast majority of humanity experiences a traffic reality different from the West
Cost-constrained environments with limited sensing

- **1.3 million** traffic-related deaths worldwide
- **90% of fatalities** on roads are from low and middle-income countries that have **48%** of world's vehicles (WHO, 2009)
  - African cities have **highest fatality rates** globally (World Bank, 2011)
- Urban areas will absorb all population growth for the next 4 decades, estimated at **2.6 billion people** (UN, 2011)
  - One in two Africans will live in a city in 2030, compared to one in three now (UN, 2005)
  - Major African cities growing at **3-5%** annually (ACID, 2010)
- High demand is coupled with lower investments in transportation and planning

Opportunistic sensing that

- Leverages explosion in cell phone adoption in growth markets
- Machine learning methods to determine emergent behavior
- Sustainable transportation recommendations

© 2013 IBM Corporation

Urbanization in Growth Markets - **Opportunities**

Doing more with less

Classical approach

- City operator, to determine how a city moves, does
  - Surveys using samples
  - Calibration using sensors
- City operates in
  - Data-rich environment
  - Resource intensive response

$$$

A data-driven future

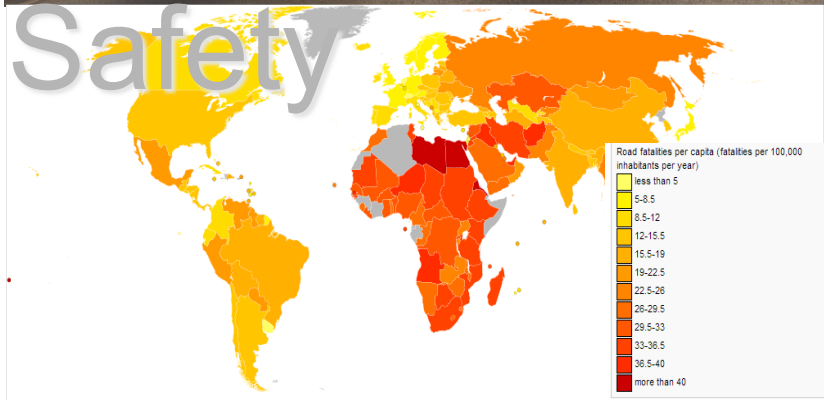
- Machine driven demand estimates from telecommunication data
  - robust to uncertainties in mobility data
  - Large samples
- Optimization of public transport services as a sustainable mode to adapt to citizen needs

© 2013 IBM Corporation

Abidjan, Côte d'Ivoire

Transport

- **439 buses** for formal operator with 17 express and 67 routes for **5 million residents**
- 5,000 taxis and 11,000 mini-buses for informal operators
- Problems
  - **Rigidity** in routes and terminals with little regard for passenger demand
  - Poor **safety** record
  - High **congestion**, with mini-buses accounting for 50% of traffic along certain corridors
  - **Environmental** impact: high emissions and low fuel efficiency
- Informal transport fills the void that the operator cannot meet at lower costs but with negative outcomes



SmarterCities Technology Centre

© 2013 IBM Corporation

Orange Data 4 Development Challenge
Côte d'Ivoire

Mobile phone data

- **13.3 million cell phones** in Cote d' Ivoire (2009)
- Mobility patterns from **2.5 billion** call records (calls and SMS)
- Estimate of travel sequences between cell towers using a sample of **500,000 users**
- Data for **5 months** between 407 antennas in Abidjan

© 2013 IBM Corporation

AllAboard

AllAboard assists cities in improving their transport network in order to

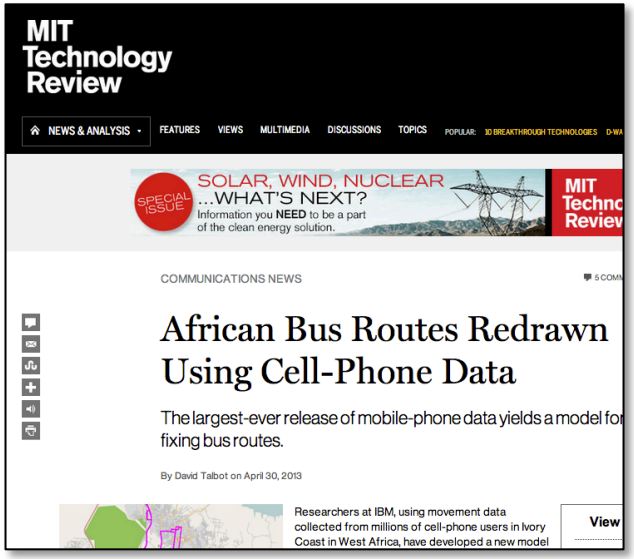
- Maximize met travel demand
- Minimize passengers' travel and wait times

The system

- relies on data from Telecom networks, potentially taking into account all cellphone users
- solves a large scale transit optimization, maximizing the quality of services provided to the travellers, on a given budget

The system allows evaluating:

- Effective ridership of roads and transit services
- Expected increased ridership and reduced travel and wait times after new services are added (e.g. new routes)
- The value for money of the proposed improvement



SmarterCities Technology Centre

© 2013 IBM Corporation

AllAboard applied to Abidjan, Ivory Coast

Input:

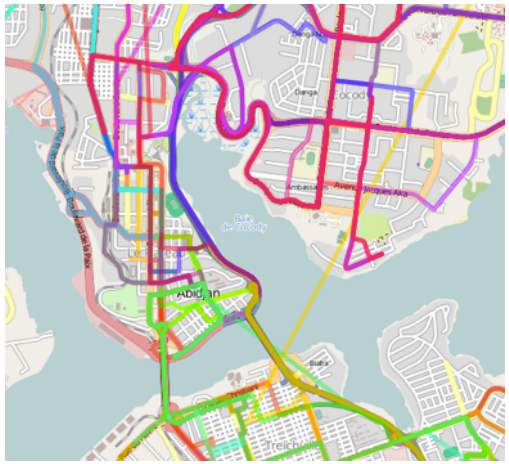
- Cellphone location data from 500,000 users in Ivory Coast[1]
- Existing transit network consisting of 17 express bus routes and 67 regular bus routes

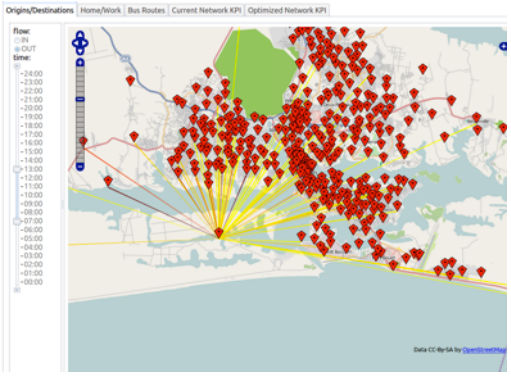
Pre-processing

- Cellphone data was able to estimate 15,000 time-varying Origin-Destination flows
- 30 frequent location sequences were extracted

Output

- Optimization model selected 4 new routes
- 22 routes had increased ridership
- Additional met demand that resulted in citywide travel time decreased by 10%



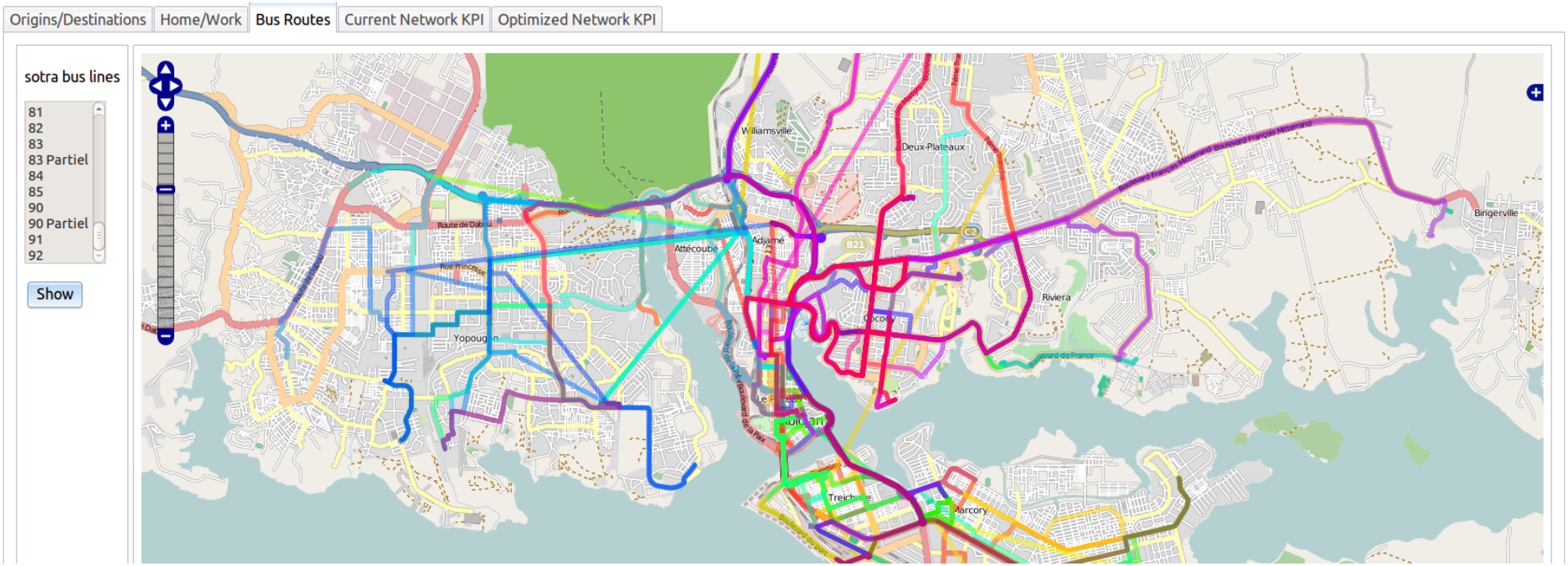


(1) Data from Orange, under the D4D Challenge

© 2013 IBM Corporation

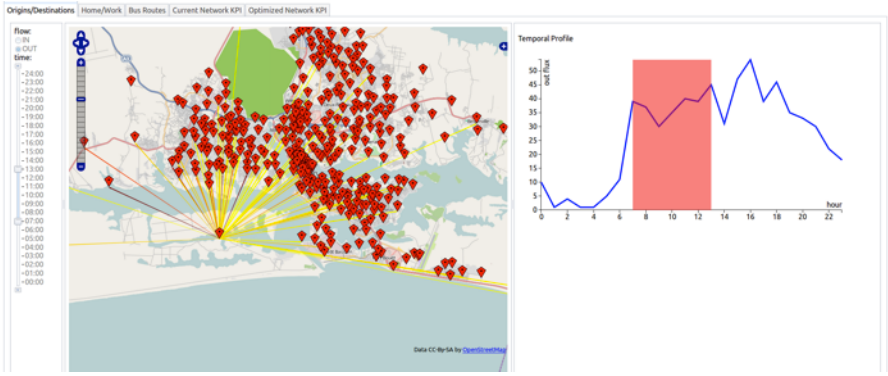
Sotra Network

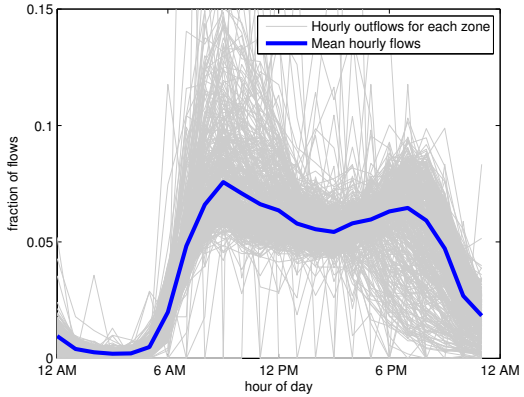
From a set of 301 distinct bus stops, 203 were geolocated using OSM + googlemap.
These stops represent 17 express bus routes, 67 regular bus routes (called the monbus),
and 1 special route meant for trader with heavy goods (called marche bus).



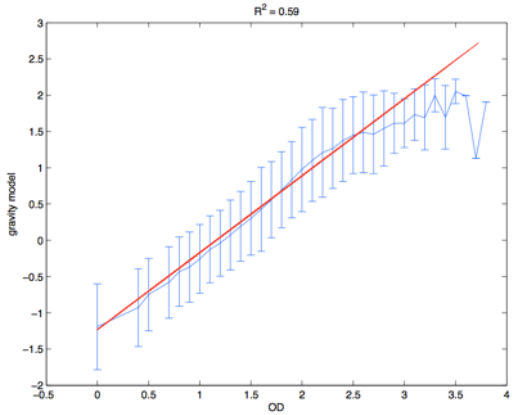
© 2013 IBM Corporation

O/D matrix estimation





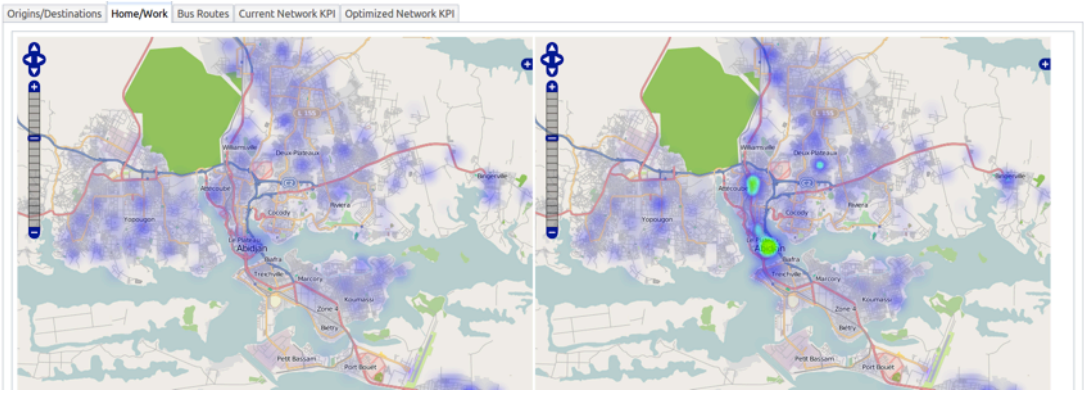
Hourly distribution of outflows



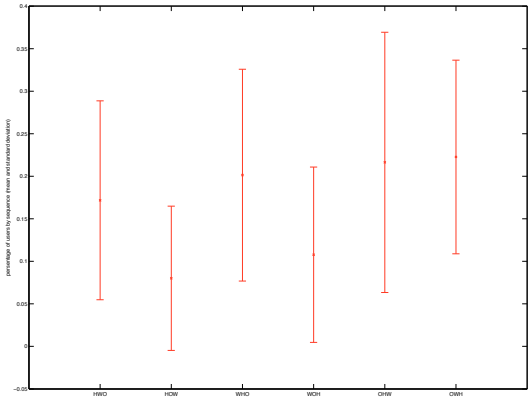
Comparison with the gravity model

© 2013 IBM Corporation

Home work location and frequent sequence labeling



For each user associated to a sequence, we label the visited stops as home (H), work (W) or other (O). (e.g. For instance strings of type ”OOO” represent the case when all 3 locations in the sequence do not correspond to either home or work cell)



80% of the cases, the sequences present both home and work

The “Other” place is visited, in most of the cases, before or after the commuting pattern H/W or W/H

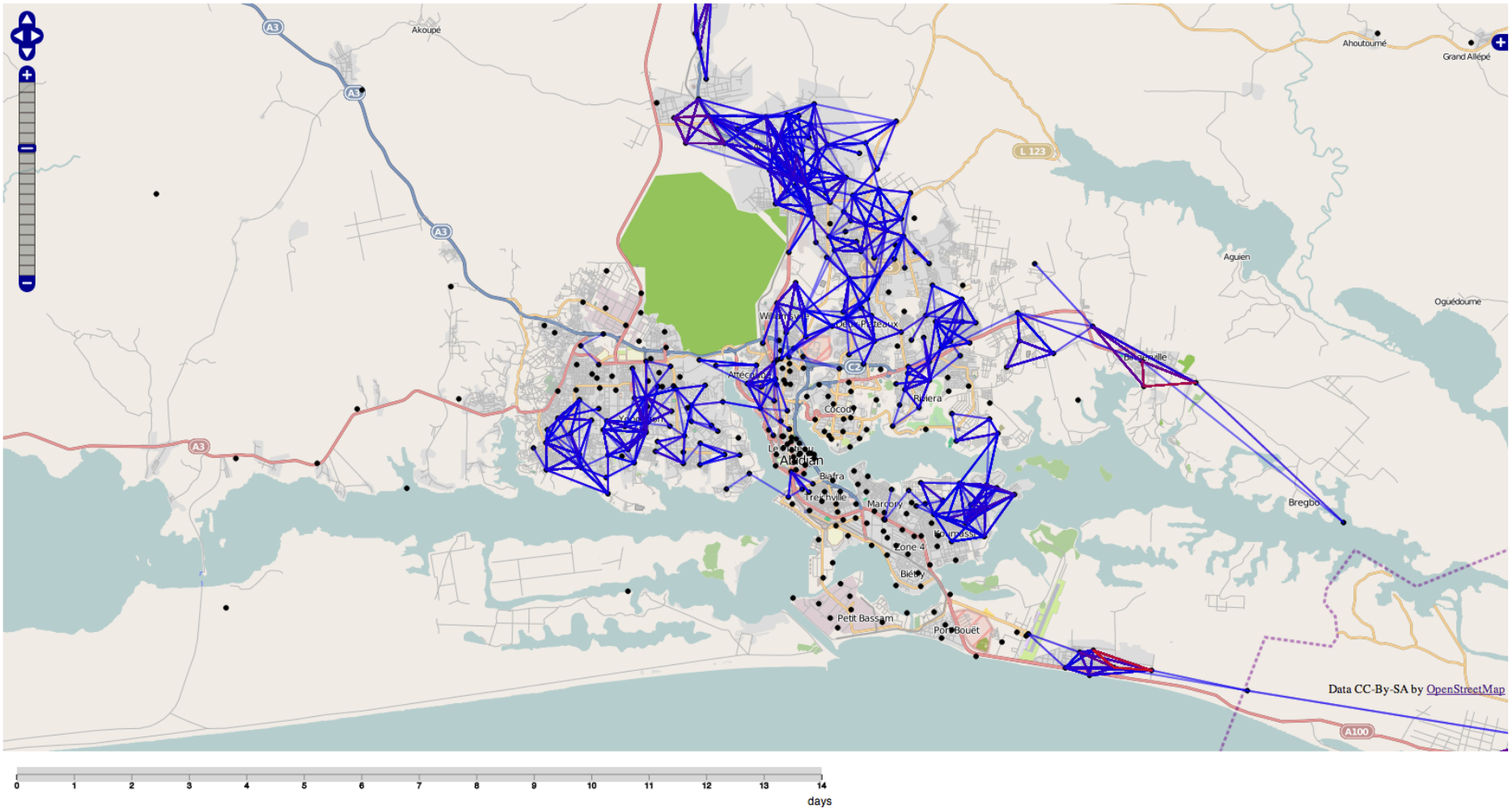
© 2013 IBM Corporation

IBM Research – Ireland - 2013

AllAboard Optimizing public transport using cellphone data

IBM Research

Antenna-to-Antenna Travel Flow | Region-to-Region Travel Flow | Home/Work Locations | Existing Transit Network | Ridership on Existing Transit Network | Travel Patterns | Ridership on Optimal Transit Network | Ridership Comparison | About

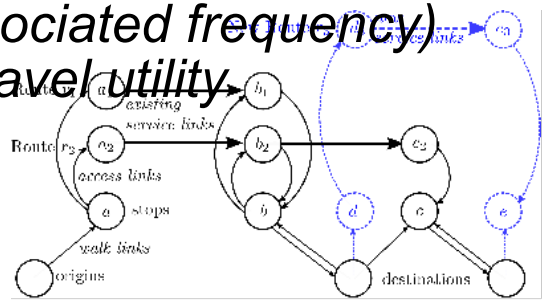


SmarterCities Technology Centre

© 2013 IBM Corporation

Optimization

- Given
  - *Existing transit network*
    - *Stops, travel times, services, frequencies*
  - *Set of frequent patterns (candidate routes)*
  - *Origin-Destination flows*
  - *Budget constraint (fleet size)*
- *Find…*
  - *Set of routes (with associated frequency) that maximizes user travel utility*



© 2013 IBM Corporation

Optimization Model (user perspective)

- Minimize travel time + waiting time
- Subject to
- Flow conservation
- Relate flows and frequencies
- New service constraints
- Budget
- Bound frequency
- Non-negative continuous variables

$$\min \sum_{k \in K} \left[\sum_{(i,j) \in A} c_{ij} v_{ijk} + \sum_{i \in V} w_{ik} \right]$$

subject to,

$$\sum_{j:(i,j) \in A} v_{ijk} - \sum_{j:(j,i) \in A} v_{ijk} = g_{ik} \forall i \in V, k \in K$$

$$v_{ijk} - q_{ij} w_{ik} \leq 0 \qquad \forall i \in V', k \in K$$

nonlinear

$$v_{ijk} - \sum_{r \in R} \delta_{ijr} f_r w_{ik} \leq 0 \qquad \forall (i,j) \in NA, k \in K$$

$$\sum_{r \in R} \frac{t_r}{60} f_r \leq \bar{C}$$

$$0 \leq f_r \leq u_r \qquad \forall r \in R$$

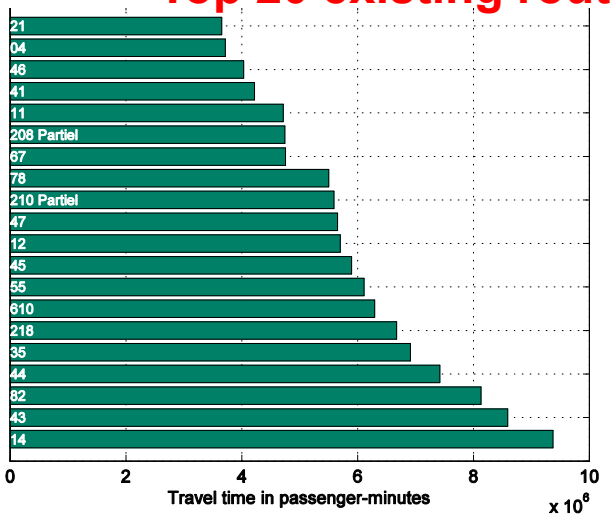
$$v_{ijk}, w_{ik}, f_r. \in \mathbb{R}^+$$

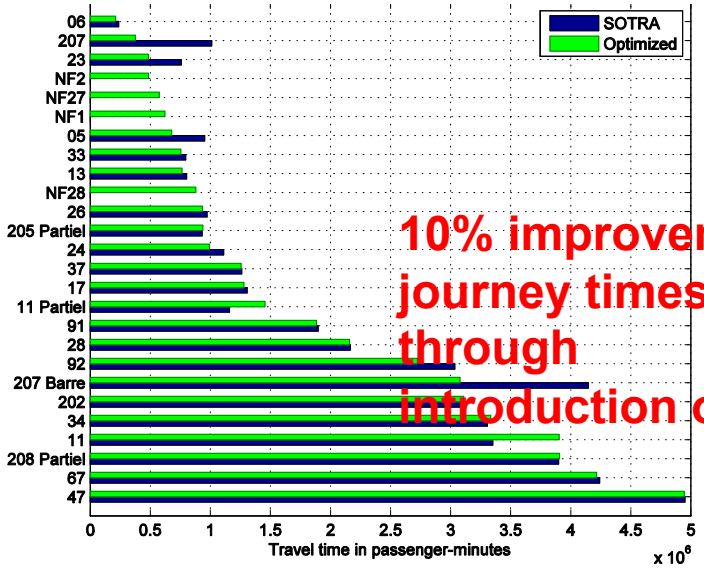
© 2013 IBM Corporation

Implementation and results

- 15,000 OD pairs
- 92 existing transit routes
- 30 candidate frequent services + 35 “shortest path” candidates
- Separable approximation model in CPLEX

Top 20 existing routes by flow





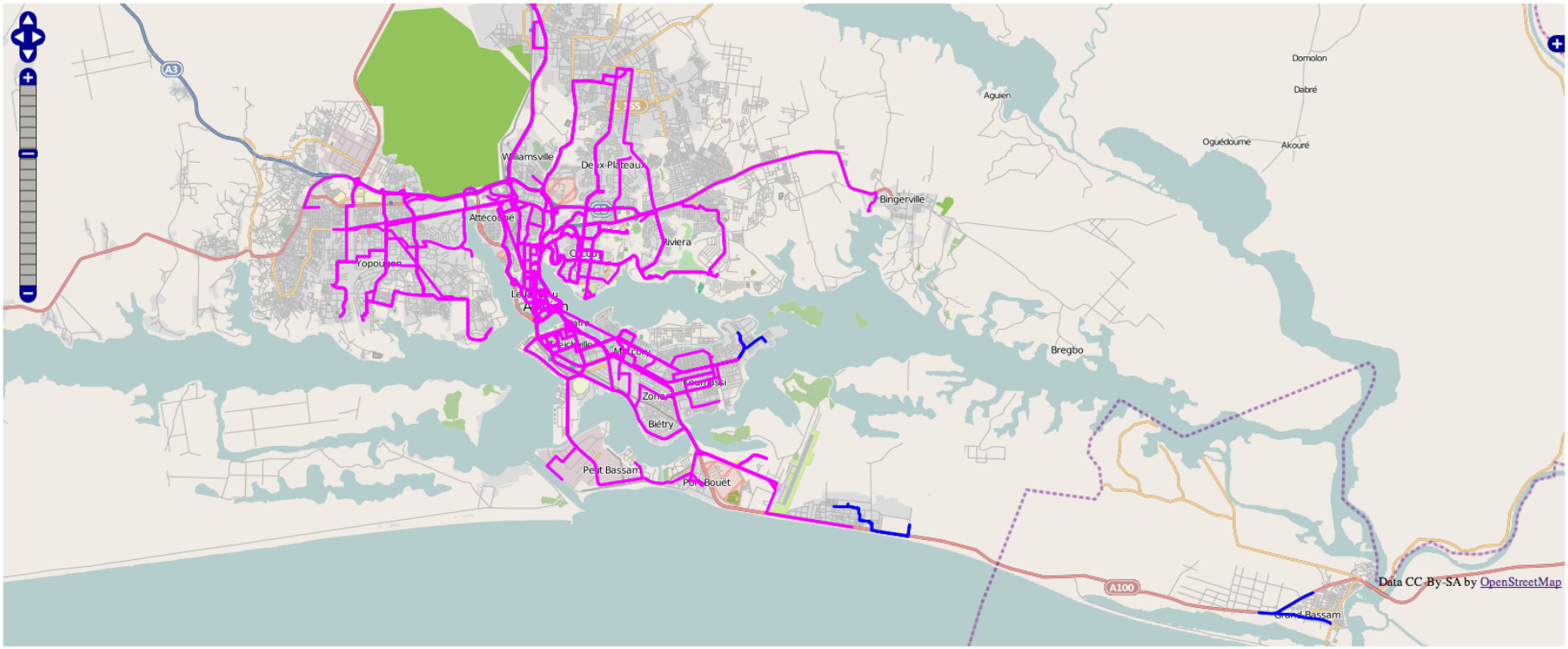
10% improvement in journey times system-wide through introduction of 4 routes

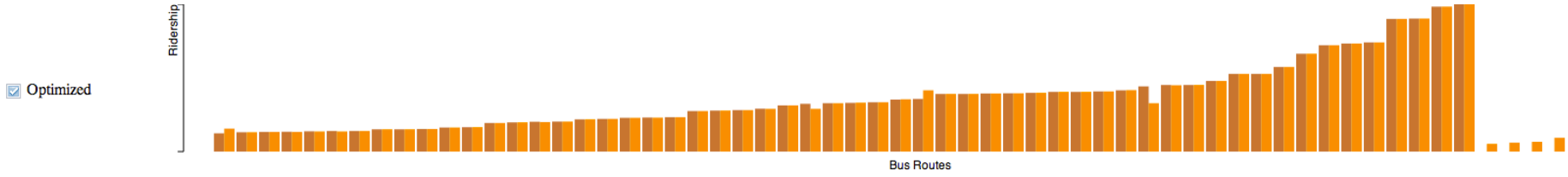
© 2013 IBM Corporation

AllAboard Optimizing public transport using cellphone data

IBM Research

Antenna-to-Antenna Travel Flow | Region-to-Region Travel Flow | Home/Work Locations | Existing Transit Network | Ridership on Existing Transit Network | Travel Patterns | Ridership on Optimal Transit Network | Ridership Comparison | About





© 2013 IBM Corporation