Sizing, Incentives and Redistribution in Bike-sharing Systems

Nicolas Gast

G-scop seminar, dec 2011, Grenoble

1. joint work with Christine Fricker (Inria)
Outline

1. Introduction and model
2. Detailed study of the homogeneous case
3. Adding some Heterogeneity
4. Conclusion and future work
A new transportation system.

- Bike sharing systems started in the 60s.
- > 400 cities. Ex: Lausanne, Barcelona, Montreal, Washington.
- Various size: from 200 to more than 50,000 bikes.

Example of Velib’:
- 20,000 bikes
- 2,000 stations.

Usage:
- Take a bike from any station.
- Use it.
- Return it to a station of your choice.

Map of Velib’ stations in Paris (France).
Public but different from public transportation

Business model (in most of the cities)
- publicity in exchange of guarantee of service.

Many advantages:
- Good for the town (pollution, traffic jams, health, image);
- Good for the citizen (cheap, quick, no bike to buy, no risk of theft).

However: congestions problems.

Empty station  Full station  Good stations

:-(  :(  :)

Problematic stations

- Goal of city: minimize the number of problematic stations.
- Goal of operator: minimize the running cost.
How to manage them?

Identify bottlenecks:

- **time dependent arrival rate**: daily period
- **heterogeneity**: popular or non-popular stations (housing and working areas, uphill and downhill stations,...)
- **random choices** of users.

**Strategic decisions**

- **Planning**: number of stations, location, size.
- **Long term operation decisions**: static pricing, number of bikes.
- **Short term operating decisions**: dynamic pricing, repositioning.

**Research challenges**:

- **Quantify** what can be asked by the city.
- **Modelling**: temporal and spatial dependencies.
Our approach

Congestion due to flows and random choices

In this talk : study the impact of random choices

1. Qualitative behavior and quantitative impact of different factors.
2. Strategies: redistribution (trucks) and incentives (pricing).

Related work:

- Traces analysis, clustering (Borgnat et al. 10, Vogel et al. 11, Nair et al. 11)
- Redistribution based of forecast [Raviv et al. 11, Chemla et al. 09]
- Few stochastic models. In a similar context : limiting regime with infinite capacity [Malyshev Yakovlev 96, Georges Xia 10]
Outline

1. Introduction and model

2. Detailed study of the homogeneous case

3. Adding some Heterogeneity

4. Conclusion and future work
The simplest case: homogeneous

\[
C = 4
\]

For all \( N \) stations:
- Fixed capacity \( C \)

Will be extended to non-homogeneous:
- arrival rate, routing probability
The simplest case: homogeneous

For all \( N \) stations:
- Fixed capacity \( C \)
- Arrival rate \( \lambda \).

Will be extended to non-homogeneous:
- Arrival rate, routing probability
The simplest case: homogeneous

For all $N$ stations:
- Fixed capacity $C$
- Arrival rate $\lambda$
- Routing matrix: homogeneous.
- Travel time: exponential of mean $1/\mu$.

Will be extended to non-homogeneous:
- Arrival rate, routing probability
The simplest case: homogeneous

For all $N$ stations:
- Fixed capacity $C$
- Arrival rate $\lambda$. 
- Routing matrix: homogeneous.
- Travel time: exponential of mean $1/\mu$.
- Other destination chosen if full ($\approx$ local search).

Will be extended to non-homogeneous:
- arrival rate, routing probability
A first result: steady state distribution of stations

- Compute the fraction of station with \( i \) bikes.

**Theorem**

There exists \( \rho \), such that in steady state, as \( N \) goes to infinity:

\[
    x_i = \frac{1}{N} \# \{ \text{stations with } i \text{ bikes} \} \propto \rho^i.
\]

We have \( \rho \leq 1 \) iff \( s \leq \frac{C}{2} + \frac{\lambda}{\mu} \) where \( s \) be the average number of bikes per stations.

\[
    s < \frac{C}{2} + \frac{\lambda}{\mu}
\]

\[
    s = \frac{C}{2} + \frac{\lambda}{\mu}
\]

\[
    s > \frac{C}{2} + \frac{\lambda}{\mu}
\]
Proof based on mean field approximation

\[ x_i = \frac{1}{N} \# \{ \text{stations with } i \text{ bikes} \} \]

For fixed \( N \), \( X_i \) is a complicated stochastic process

- Reversible process but steady state not explicit.
Proof based on mean field approximation

\[ x_i = \frac{1}{N} \# \{ \text{stations with i bikes} \} \propto \rho^i \]

For fixed \( N \), \( X_i \) is a complicated stochastic process

- Reversible process but steady state not explicit.

Use mean field approximation [Kurtz 79]

- Study the system when the number of stations \( N \) goes to infinity.

System described by an ODE

- The ODE has a unique fixed point.
- Closed-form formula.
Consequences: optimal performance for \( s \approx \frac{C}{2} \)

Fraction of problematic stations (=empty+full) \( x_0 + x_C \) is minimal for

\[
\rho = 1 \quad \text{i.e.} \quad s = s_c \overset{\text{def}}{=} \frac{\lambda}{\mu} + \frac{C}{2}
\]

- Prop. of problematic stations is at least \( \frac{2}{(C + 1)} \) and “flat” at \( s_c \).

Ex: for \( C = 30 \): at least 6.5% of problematic stations.

\( y \)-axis: Prop. of problematic stations. \( x \)-axis: number of bikes/station \( s \).
Improvement by dynamic pricing: “two choices” rule

- Users can observe the occupation of stations.
- Users choose the least loaded among 2 stations close to destination to return the bike (ex: force by pricing)
Improvement by dynamic pricing: “two choices” rule

- Users can observe the occupation of stations.
- Users choose the least loaded among 2 stations close to destination to return the bike (ex: force by pricing)

Paradigm known as “the power of two choices”:
- Comes from balls and bills [Azar et al. 94]
- Drastic improvement of service time in server farm [Vvedenskaya 96, Mitzenmacher 96]

Question: what is the effect on bike-sharing systems?

Characteristics:

1. Finite capacity of stations.
2. Strong geometry: choice among neighbors.
Two choices – finite capacity but no geometry

With no geometry, we can solve in close-form.

- Proof uses similar mean field argument.

Choosing two stations at random, improves perf. from $\frac{1}{C}$ to $2^{-C}$
Two choices – taking geometry into account

Problem hard to solve: mean field do not apply (geometry) :(.

- Existing results for balls and bins (see [Kenthapadi et al. 06])
- Only numerical results exists for server farms (ex: [Mitzenmacher 96])

We rely on simulation

Occupancy of stations
x-axis = occupation of station.
y-axis: proportion of stations.

Recall: with no incentives, the distribution would be uniform.

Empirically:

- with geometry 2D: proportion of problematic stations is \( \approx 2^{-C/2} \).
  (recall: with no-geometry: \( 2^{-C} \), with no incentive: \( 1/C \)).
Improvement by redistribution

Same model as before with a truck

Question: what should $\gamma$ be? 10%, 20%, more?
Improvement by redistribution

Same model as before with a truck

With rate $\gamma \cdot \lambda$:
- Take a bike from the most loaded.
- Put it in the least loaded.

Question: what should $\gamma$ be? 10%, 20%, more?

Introduction and model
Homogeneous case
Heterogeneous case
Conclusion and future work
**Improvement by redistribution**

\[ C = 4 \]

\[ C = 4 \]

\[ C = 4 \]

\[ C = 4 \]

\[ \gamma \cdot \lambda \]

Same model as before with a truck

With rate \( \gamma \cdot \lambda \):

- Take a bike from the most loaded.
- Put it in the least loaded.

**Question**: what should \( \gamma \) be? 10\%, 20\%, more?
Improvement by redistribution

Same model as before with a truck

With rate $\gamma \cdot \lambda$:
- Take a bike from the most loaded.
- Put it in the least loaded.

Question: what should $\gamma$ be? 10%, 20%, more?
Optimal rate of regulation is \( \frac{1}{(C - 1)} \)

Recall \( C \) is the capacity, \( s \) the fleet size and \( N \) the number of stations.

### Theorem

As \( N \) goes to infinity, we have:

- The number of problematic stations decreases as \( \gamma \) increases.
- If \( \gamma > \frac{1}{2[C-(s-\lambda/\mu)]-1} \), then there is no problematic stations.

For example: if \( s = \frac{C}{2} + \frac{\lambda}{\mu} \), a regulation rate of \( \frac{1}{(C - 1)} \) suffices.

### Proof

Again mean field approximation but with discontinuous dynamics

- The dynamical system is described by a differential inclusion

\[ \dot{x} \in F(x). \]

- The DI has a unique solution. We can solve in close-form.

See [Gast Gaujal 2010].
Optimal rate of regulation, illustration

Example: capacity is $C = 10$. Fleet size is 3, 5 or 7 bikes/stations.

1. No regulation, $\gamma = 0$

2. Regulation ($\gamma = 10\%$).

$x$-axis = occupancy of stations, from 0 to 10.
$y$-axis = proportion of stations.
## Conclusion on the homogeneous model

<table>
<thead>
<tr>
<th></th>
<th>prop. of problematic stations</th>
<th>ex: ( N = 30 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original model</td>
<td>( \frac{1}{C} )</td>
<td>6.5%</td>
</tr>
<tr>
<td>Two choices</td>
<td>( \begin{align*} \frac{1}{C} \ \frac{1}{C/2} \end{align*} )</td>
<td>( 10^{-9} \approx 0 ) ( 10^{-4.5} )</td>
</tr>
<tr>
<td>Regulation</td>
<td>( \gamma &gt; \frac{1}{C-1} )</td>
<td>0</td>
</tr>
</tbody>
</table>

### However

As mentioned before, there are some important factors:

- **time dependent arrival rate**: daily period
- **heterogeneity**: popular or non popular stations
  (housing and working areas, uphill and downhill stations,...)
Outline

1. Introduction and model
2. Detailed study of the homogeneous case
3. Adding some Heterogeneity
4. Conclusion and future work
Heterogeneous model

For each station $i$:
- Fixed capacity $C_i$

$C_1 = 5$

$C_2 = 3$

$C_3 = 4$
Heterogeneous model

For each station $i$:
- Fixed capacity $C_i$
- Arrival rate $\lambda_i$.

$$C_1 = 5$$
$$C_2 = 3$$
$$C_3 = 4$$

For each station $i$:
- Fixed capacity $C_i$
- Arrival rate $\lambda_i$. 

Travel time: exponential of mean $1/\mu$. 

Local search if full.
Heterogeneous model

For each station $i$:
- Fixed capacity $C_i$
- Arrival rate $\lambda_i$.
- Popularity of station $p_i$.
- Travel time: exponential of mean $1/\mu$.
- Local search if full.
Steady state performance

There are $N$ stations. Assume that as $N$ goes to infinity, the popularity of the parameters $p_i = (\lambda_i, p_i)$ goes to some distribution.

**Theorem (Propagation of chaos-like result)**

There exists a function $\rho(p)$ such that for all $k$, if stations $1, \ldots, k$ have parameter $p_1, \ldots, p_k$, then, as $N$ goes to infinity:

$$P(\#\{\text{bikes in stations } j\} = i_j \text{ for } j = 1..k) \propto \prod_{j=1}^{k} \rho(p_j)^{i_j}$$

Depending on popularity, stations have different behaviors:

- **Popular start**
- **→**
- **Popular destination**
Steady-state performance: numerical example

- In general, $\rho$ is the solution of a fixed-point equation.
- Can be plotted in closed form for particular cases.

**Figure:** Two types of stations: popular and non-popular for arrivals: $\lambda_1/\lambda_2 = 2$. 

**Introduction and model**

**Homogeneous case**

**Heterogeneous case**

**Conclusion and future work**
Outline

1. Introduction and model
2. Detailed study of the homogeneous case
3. Adding some Heterogeneity
4. Conclusion and future work
Good understanding of the symmetric model

- **Performance poor**: $1/C$ problematic stations (even for symmetric!).
- Simple incentives helps a lot: $2^{-C/2}$.
- **Optimal regulation** rate is function of capacity: $1/C$.

Current and future work

- Building a **realistic model** of Paris (using traces).
- Analyze **transient and steady-state** behavior.
- Difference effect of **flows vs random perturbations**.
- Develop models to **approximate the influence of geometry**.