# An Efficient Content Delivery Infrastructure Leveraging the Public Transportation Network

Qiankun Su, Katia Jaffrès-Runser, Gentian Jakllari and Charly Poulliat

IRIT, INP-ENSEEIHT, University of Toulouse

GDR RSD - Journées non thématiques
Sophia-Antipolis
12 janvier 2017

# Context

## Growing urbanization

- 70% of the world population will live in urban areas by 2050[1].
- Increased demand for public transportation.

## Growing mobile traffic

- The mobile data traffic will be increased nearly 8-fold between 2015 and 2020[2].
- 5G deployments are not expected until at least 2020[2].

[1] **un2014**.

[2] **cisco2016global**.

# Main idea and contribution

## Main idea

A significant part of mobile content is consumed while people use public transportation

- Leverage public transportation network (PTN) to offload mobile traffic of bus customers from cellular network

## Main contribution

A novel content delivery infrastructure where

- buses act as data mules, interconnected by wireless access points (APs) at selected bus stations.
- push up to 1TByte within 12 hours in Paris [3]

[3] AP rate: 150Mb/s

# Outline

A content delivery infrastructure using PTNs

XOR network coding for PTNs

Towards a cost-effective design

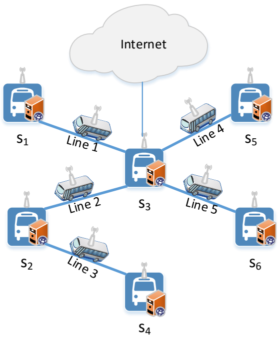
# Scenario description

Public buses act as data mules, creating a delay tolerant network (DTN)

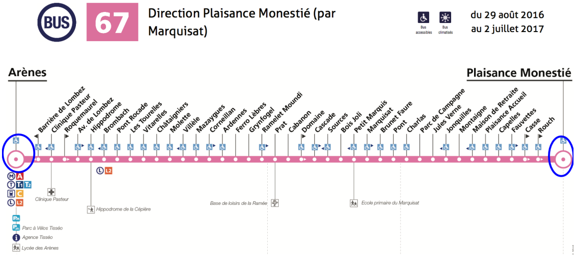
- Contents are obtained from nodes connected to the Internet
- Passengers download/upload contents from/to buses

Previous solutions

- UMass DieselNet: testbed for DTN routing (MaxProp).
- DakNet: low-cost Internet to remote villages.



# Where to install WiFi APs?

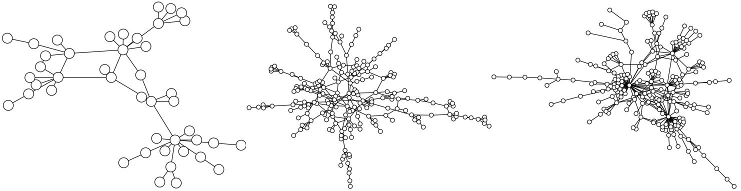



- The waiting time of buses at intermediate stops is very short.
  - Inter-station travel time $\leq$ 4 minutes in Paris
  - Waiting time at end stations : 60% over 10 minutes.
- APs can only be deployed at end stations:
  - Larger bandwidth at interconnection / reduced deployment.
  - Several bus lines cross at end stations.

# Modeling

Our infrastructure can be modeled as an undirected graph where

- nodes represent bus end stations
- edges represent bus lines



(a) Toulouse (b) Paris (c) Helsinki

Fig. 1: The biggest connected component of public transportation networks.

# Our routing policy (1/2)

Main routing protocols in DTN are designed for

- non-predictable mobility patterns

The features of DTNs created by PTNs,

- The network topology is stable.
- The behavior of buses is predictable.

Our routing policy,

- messages are delivered following the shortest path
- pre-calculate routing tables for each end station

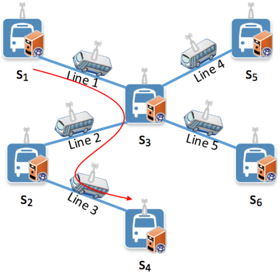



Fig. 2: Content delivery using PTNs.

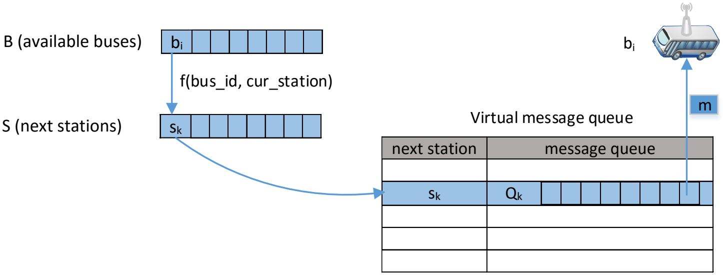
# Our routing policy (2/2)

Receive a message $m$ at an end station

- extract $m$'s destination, look up its next-hop stations $s_k$
- $m$ is placed into $Q_k$ that stores messages going to $s_k$

Send a message $m$ at an end station

- $B$: a list of buses currently waiting at the station
- $S$: a corresponding list of next-hop stations



# Outline

# Problem statement and motivation

- PTNs are built around the concept of hubs with many bus lines.
- AP: fair medium access control.

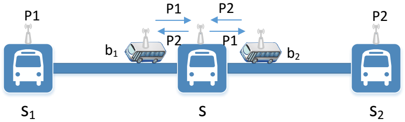



Fig. 3: Light network load

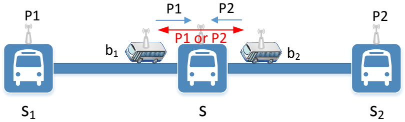



Fig. 4: Heavy network load

$\Rightarrow$ Such an imbalance results in a significant drop in throughput under heavy traffic conditions.

# XOR network coding

XOR network coding implementation

- pairwise inter-session flows[4]
- hop-by-hop

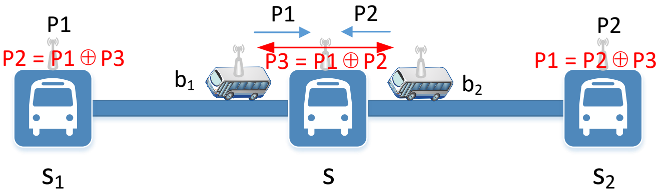



Fig. 5: The benefits of XOR network coding.

[4] **ICCC15**.

# XOR network coding implementation

Encoding procedures:

- Message queues $Q_{ij}$ are indexed by the previous station $s_i$ and the next station $s_j$ of messages.

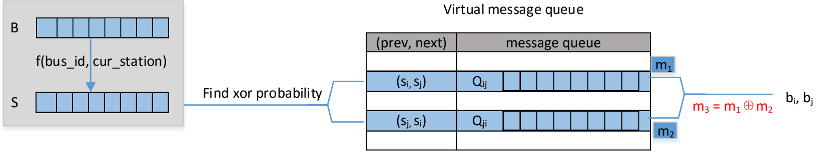



Fig. 6: Encoding procedures

Decoding procedures:

- receive a xor-ed message $m = m_i \oplus m_j$
- xor-ing again with the message previously sent, $m_j = m \oplus m_i$

# Simulation setup

We leave for further investigations how content is requested, updated and fetched. The goal is to show the pure network coding benefits in our infrastructure.

## Simulation settings

- the ONE (Opportunistic Network Environment simulator)
- Bus schedules [5] for Toulouse, Paris, Helsinki, 7am - 7pm

## Data flows (multiple unicast flows)

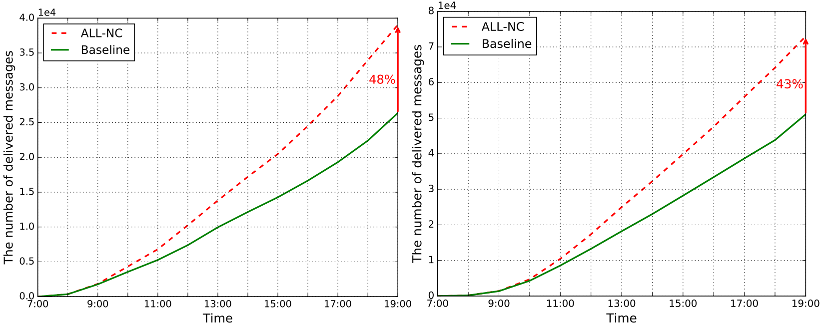
- A message is created at every end station with a given creation period ($\Delta = 20$ *seconds*).
- The message destination is selected uniformly at random among all the stations.

[5] using GTFS (General Transit Feed Specification), developed by Google

# Benefits of network coding

## Cumulative number of delivered messages

- ALL-NC: with network coding
- Baseline: without network coding




(a) Toulouse (b) Paris

# Outline

# Problem statement

Network coding is really beneficial for our infrastructure. However, we want to minimize the deployment of APs performing network coding for several reasons:

- For APs to perform fast network coding operations, we may need a more complex and powerful architecture.
- Network coding is threatened by specific attacks, and a single corrupted flow may be detrimental to several others.

# 3-Tier architecture

## Divide stations into 3 tiers

- 1-Tier: No AP
- 2-Tier: Regular AP
- 3-Tier: Network coding enabled AP

The goal is to:

- maximize the number of delivered messages
- minimize the cost of deployment

# 2-Tier node selection

A minimum connected dominating sets,

- minimize the number of wireless AP
- guarantee the end to end connectivity

A CDS is formed by *M Rai et al.*[6].

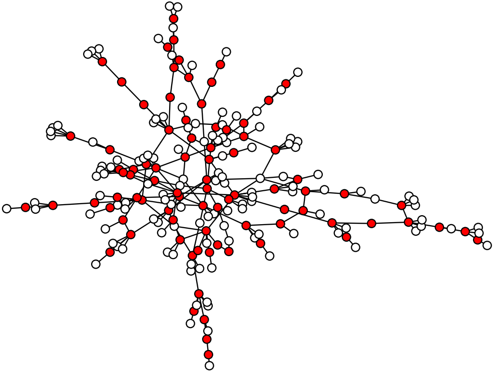


Fig. 7: A CDS in Paris topology (in red).

[6] **rai2009new**.

# 2-Tier, the decrease of APs

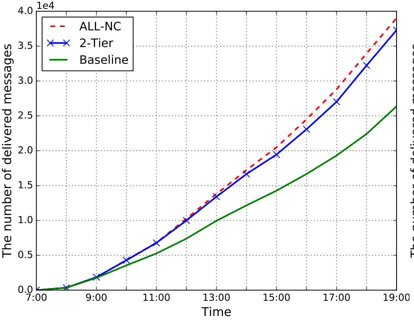
Save around three times of wireless APs.

| City | Baseline | ALL-NC | 2-Tier |
|---|---|---|---|
| Toulouse | 44 | 44 | 13 |
| Paris | 213 | 213 | 85 |
| Helsinki | 217 | 217 | 60 |

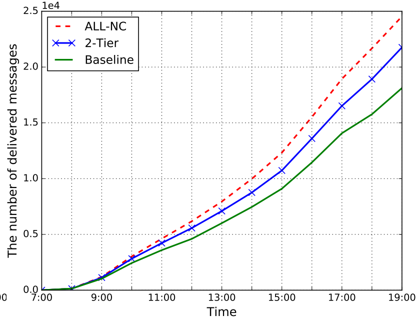
Table 1: The 2-Tier architecture reduces the required number of interfaces by approximately a factor of 3.

# 2-Tier, performance evaluation

- 2-Tier: stations belong to CDS equipped with network coding enabled APs




(a) Toulouse




(b) Helsinki

Fig. 8: Number of messages delivered for Baseline, ALL-NC and 2-Tier.

# 3-Tier

Select the top $n$ nodes from CDS to install network coding AP

- Large benefits of network coding if existing a lot of cross flows
- Identify nodes with high degree, betweenness, PageRank

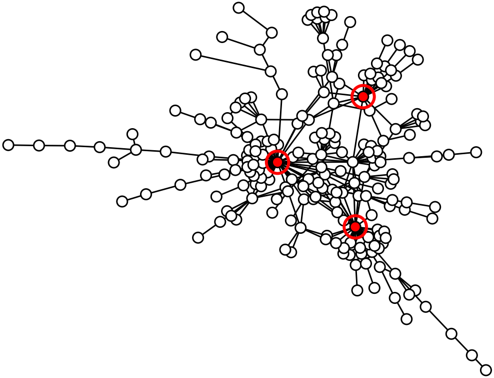
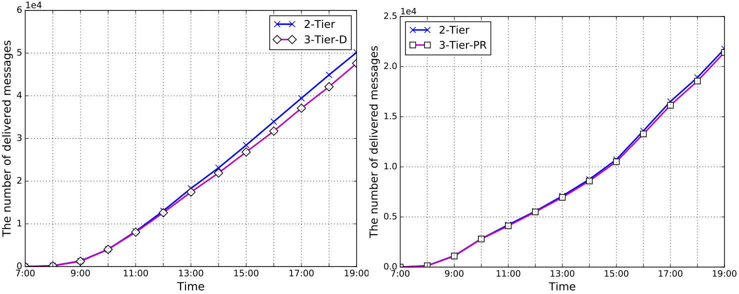


Fig. 9: The top 3 highest PageRank in Helsinki topology (in red)

# 3-Tier, performance evaluation

| City | 2-Tier | 3-Tier | Metric |
|---|---|---|---|
| Toulouse | 13 | 2 | Degree |
| Paris | 85 | 10 | Degree |
| Helsinki | 60 | 3 | PageRank |

Table 2: 3-Tier reduces the number of such interfaces by over an order of magnitude.




(a) Paris

(b) Helsinki

Fig. 10: Packets delivered for 2-Tier and 3-Tier.

# Cost-effectiveness analysis

The cost of

- a regular wireless AP: 1
- a secure network coding enabled AP: $C$ ($C > 1$)
- Y axis: the cost effectiveness $= \frac{The\ number\ of\ delivered\ messages}{The\ deployment\ cost}$

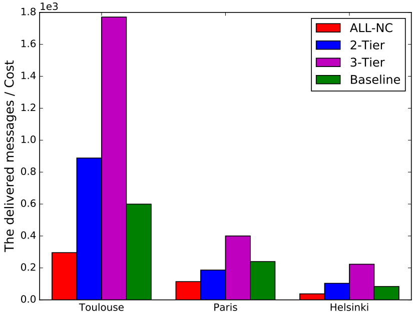



Fig. 11: The cost effectiveness for all architectures ($C = 3$).

# Conclusion and perspectives

## Contribution

A cost-effective content delivery infrastructure that

- offloads a large amount of data, e.g., 1 TByte within 12 hours in Paris topology thanks to a careful deployment of network coding enabled APs,
- reduces the number of wireless APs by a factor of 3

## Perspectives

- Analyze & design a centrality metric for simple selection of network coding locations,
- Capture realistic delay-tolerant traffic patterns (content delivery infrastructure, private cloud, ?)
- Design learning mechanism to push popular content to the buses for end users, etc.

# Thank you for your attention.

Questions ?