Collaborative Data Dissemination in Opportunistic Vehicular Networks

Yong Li[a], Zhaocheng Wang[a], Depeng Jin[a], Lieguang Zeng[a], Sheng Chen[b,c]

[a] Tsinghua National Laboratory for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

[b] Electronics and Computer Science, Faculty of Physical and Applied Sciences, University of Southampton, Southampton SO17 1BJ, UK
sqc@ecs.soton.ac.uk

[c] Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia

Progress In Electromagnetics Research Symposium
Kuala Lumpur, Malaysia, March 27-30, 2012

Outline

1. Introduction
   - Motivations
   - Our Approach
2. Problem Formulation
   - System Definition
   - Dissemination Algorithms
3. Simulation Results
   - Experimental Setup
   - Experimental Results
4. Conclusions
   - Concluding Remarks

Outline

Cisco Forecast 2011-2016

- Global mobile data traffic grew 2.3 fold in 2011
- Mobile video traffic exceeded 50 percent for the first time in 2011
- Global mobile data traffic will increase 18-fold between 2011 and 2016
- Monthly global mobile data traffic will surpass 10 exabytes in 2016
- Two-thirds of the world's mobile data traffic will be video by 2016
- China will exceed 10 percent of global mobile data traffic in 2016

1 exabyte $= 10^{18}$ bytes

Solutions

To cope with explosive increase in traffic demands

- Upgrade to 4G (Cisco: in 2016, 4G will be 6 percent of connections, but 36 percent of total traffic)
  - **Very expensive**, even all mobile networks are 4G, soon demands will outstrip capacity
- Offload mobile data from overloaded mobile networks to fixed networks, such as WiFi
  - Many service providers are actively considering this option (Cisco: by 2016, over 3.1 exabytes of mobile data traffic will be offloaded to the fixed network by means of dual-mode devices and femtocells each month) **insufficient**
- Need looking for new offloading paradigm
  - Collaborative mobile data dissemination via **opportunistic** vehicular networks

Outline

Challenges

- Increasing number of vehicles have device-to-device communication capacities ⇒ Inter-vehicular ad hoc network
  - FCC has allocated spectrum, IEEE is working on standard
- Vehicular network is highly mobile and sometimes sparse, contacts may be short ⇒ cannot use offloading method for WiFi
- **Opportunistic contact** between vehicles offers high bandwidth communication capacity for content dissemination
- This application is extremely challenging
  1. Network contains heterogeneous vehicles/users, in terms of data preference
  2. Data items are multi-types of different delay sensitivities and sizes
  3. Data dissemination participants' storages are limited in size

Our contribution

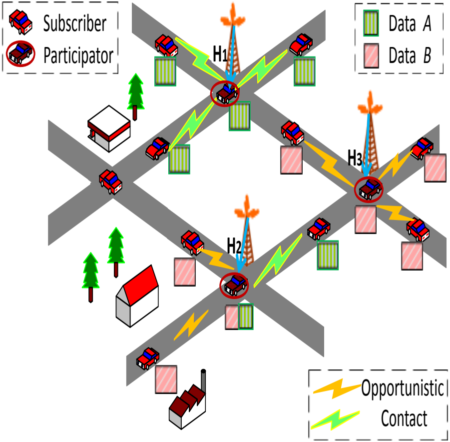
- We study collaborative data dissemination in these realistic opportunistic vehicular networks
- Our contribution is threefold:
  1. Formulate optimal data dissemination with heterogeneous data items and vehicles of limited storage as a submodular function maximisation with linear constraints
  2. Propose a heuristic and effective algorithm to solve this NP-hard problem and derive performance bound of this algorithm
  3. Demonstrate effectiveness of our algorithm in challenging opportunistic vehicular network environments through real trace-driven simulations

Outline

System Model

Service provider transmits the data to helpers/participators, who then propagate the data to other subscribers by device-to-device opportunistic communication

1. Set of helpers $\mathcal{H}$ with size $|\mathcal{H}| = H$; any $s \in \mathcal{H}$ can only offer L_s buffer size
2. Set of subscribers $\mathcal{N}$ with size $|\mathcal{N}| = N$
3. Set of data items $\mathcal{I}$ with size $|\mathcal{I}| = I$; for any $k \in \mathcal{I}$, its data length is l_k
4. Contact between vehicles i and j obeys the Poisson process with contact rate $\gamma_{i,j}$



System Utility

- Each subscriber $i \in \mathcal{N}$ is associated with an **interest** vector

$$\mathbf{w}_i = [w_{i,1} \; w_{i,2} \cdots w_{i,I}]^{\mathrm{T}}$$

where $w_{i,k}$ defines subscriber i's interest to data item $k \in \mathcal{I}$

- Let $\mathbf{X} = (x_{s,k})$, $s \in \mathcal{H}$, $k \in \mathcal{I}$, be the **storage allocation** policy, in which $x_{s,k} \in \{0, 1\}$ and $x_{s,k} = 1$ indicates that helper s stores item k in its buffer
- A lifetime T_k is assigned to each item $k \in \mathcal{I}$, and let $d_{i,k}$ be the probability that user i receives data k before deadline T_k
- Since we should maximise the expectation of the **disseminated data size** for all subscribers, objective function is

$$U(\mathbf{X}) = \sum_{k \in \mathcal{I}} l_k \sum_{i \in \mathcal{N}} d_{i,k}$$

Optimisation Problem

- Noting Poisson contact rates, system utility can be expressed as

$$U(\mathbf{X}) = \sum_{k \in \mathcal{I}} I_k \sum_{i \in \mathcal{N}} \left(1 - e^{-w_{i,k} T_k \sum_{s \in \mathcal{H}} x_{s,k} \gamma_{s,i}}\right)$$

- For subset $\mathbf{A} \subseteq \mathcal{H} \times \mathcal{I}$, we define storage allocation policy $\mathbf{X}$

$$\mathbf{X} = F(\mathbf{A}), \text{ s.t. } x_{s,k} = 1 \text{ if } (s,k) \in \mathbf{A} \text{ and } x_{s,k} = 0 \text{ if } (s,k) \notin \mathbf{A}$$

  Since $F(\mathbf{A})$ is a bijection, utility $U(\mathbf{X})$ over subset $\mathbf{A}$ is

$$\widehat{U}(\mathbf{A}) = U(F(\mathbf{A})) = \sum_{k \in \mathcal{I}} I_k \sum_{i \in \mathcal{N}} \left(1 - e^{-w_{i,k} T_k \sum_{s:(s,k) \in \mathbf{A}} \gamma_{s,i}}\right)$$

- Noting helpers' buffer constraints, we have optimisation problem

$$\max U(\mathbf{X}) \text{ or } \max \widehat{U}(\mathbf{A})$$

$$\text{s.t. } x_{s,k} \in \{0,1\}, \forall s \in \mathcal{H}, k \in \mathcal{I}, \text{ and } \sum_{k \in \mathcal{I}} I_k x_{s,k} \leq L_s, \forall s \in \mathcal{H}$$

Outline

SFM with MLCs

Optimal data dissemination is an **NP-hard** **submodular function maximisation** with **multiple linear constraints**

1. **SFM algorithm**: best available algorithm to solve this problem by approximation, but extremely high complexity
   - 5 helpers and 10 data items: complexity of just its first step of Rounding Procedure is 10^{15}
2. **Random algorithm**: each helper chooses data items randomly to fill its buffer until it is full
3. **Homogeneous algorithm**: allocates buffer assuming all helpers have same storage size and all data items have identical length
4. Our **heuristic algorithm**: a greedy-type efficient solution

Our Algorithm

- **1st greedy strategy**: select the item and helper that maximise the gain on the objective function at each stage

$$(s_0, k_0) = \arg \max_{(s,k)\in \mathbf{P}} \left(\widehat{U}(\mathbf{A} \cup (s,k)) - \widehat{U}(\mathbf{A}) \right)$$

  A: set of chosen data items and helpers, **P**: set of possible candidate solutions

- **2nd greedy strategy**: select the item and helper that maximises the per-unit-length gain on the objective function at each stage

$$(s_0, k_0) = \arg \max_{(s,k)\in \mathbf{P}} \frac{\widehat{U}(\mathbf{A} \cup (s,k)) - \widehat{U}(\mathbf{A})}{l_k}$$

- **Our heuristic algorithm** performs both these two greedy strategies and chooses the better result from the two solutions

Analysis

- **Performance bound**: the solution of our heuristic algorithm, $\mathbf{OPT}^*$, and the optimal solution of the problem, **OPT**, satisfies

$$\widehat{U}(\mathbf{OPT}^*) \geq \frac{1}{2}\left(1 - e^{-\frac{L-\mu}{L}}\right)\widehat{U}(\mathbf{OPT})$$

where $L = \sum_{s\in\mathcal{H}} L_s$ and $\mu = (H-1)\cdot \max_{k\in\mathcal{I}} l_k$

- **Algorithm complexity**: pseudo-polynomial-time algorithm with complexity $O(H^3 I^2 N)$
- **Centralised algorithm**: requires the storage sizes of vehicles and contact rates between pairs of vehicles

Since vehicles usually equip with communication interfaces for WLAN or mobile network, these global information can be communicated to service provider through the low-rate uplink channels

Outline

Vehicular Mobility Traces

1. *Shanghai* trace: records of 2,019 operational taxis during February 2007 in Shanghai city
    - A taxi sends its position report to central database by GPRS every one minute when it has passengers on board
    - but position report is sent every 15 seconds when it is vacant
    - Used in many previous investigations
2. *Beijing* trace: records of 27,000 participating taxis during May 2010 in Beijing city
    - GPS devices collect taxi locations and timestamps
    - GPRS modules report records every 15 seconds for moving taxis
    - Largest real vehicular data trace available

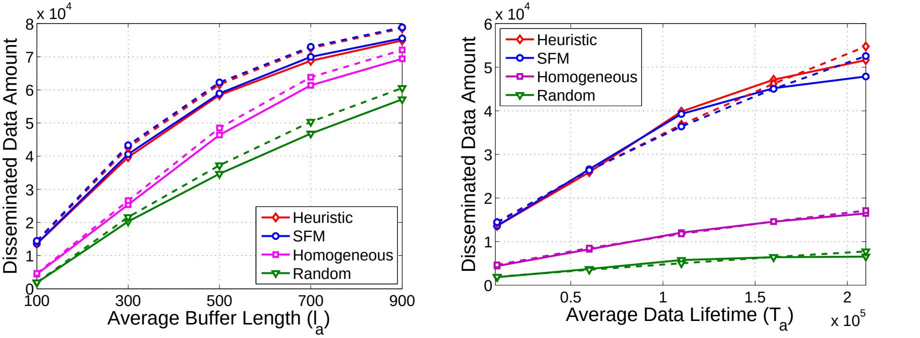
Simulation Settings

- Half of the trace is used to obtain the contact rates
- Randomly choose 10% of the vehicles as helpers and the rest as subscribers
- Number of data items is $I = 35$ for *Shanghai* trace and $I = 50$ for *Beijing* trace
- Sizes of data items are generated randomly and uniformly in $[50\,\mathrm{kB},\ 150\,\mathrm{kB}]$
- Data lifetimes follow the uniform distribution in $[0,\ 2T_a\,\mathrm{s}]$, where T_a is the average data lifetime
- Helper buffer sizes are randomly and uniformly generated in $[0,\ 2l_a\,\mathrm{kB}]$, where l_a is the average buffer size
- User interests to different data items follow the exponential distribution with an expectation of 20

Outline

Shanghai Trace

(a) fixed average data lifetime of 10000 s and variable average buffer size, and (b) fixed buffer size of 100 kB and variable average data lifetime



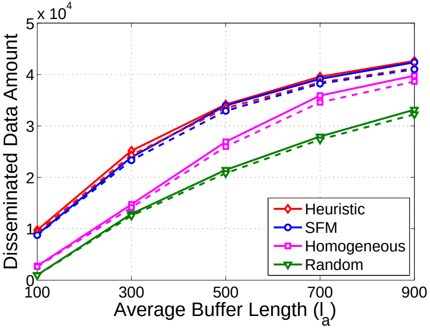
(a) (b)

Dashed curve: theoretical disseminated data size calculated by each algorithm

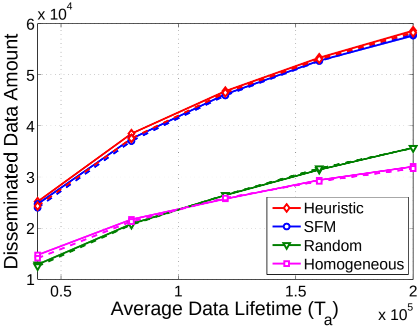
Solid curve: simulating the system with the buffer allocation strategy of each algorithm

Beijing Trace

(a) fixed average data lifetime of 40000 s and variable average buffer size, and (b) fixed buffer size of 300 kB and variable average data lifetime



(a)



(b)

Dashed curve: theoretical disseminated data size calculated by each algorithm

Solid curve: simulating the system with the buffer allocation strategy of each algorithm

Discussions

- *Shanghai* trace: Our **Heuristic Algorithm** achieves the same or slightly better performance as or than **SFM Algorithm**
- *Beijing* trace: Our **Heuristic Algorithm** achieves the same or slightly better performance as or than **SFM Algorithm**
- Average running time of SFM Algorithm is **0.53 hours**, while our Heuristic Algorithm only takes **a few seconds**
- This is significant, as SFM Algorithm is regarded as best available scheme for solving this NP-hard Problem
- Our Heuristic Algorithm offers at least the same performance as SFM Algorithm with a much lower complexity, making it practical

Outline

Summary

1. We have studied collaborative mobile data dissemination in realistic opportunistic heterogeneous vehicular networks
   - Multiple data items with different delay sensitivities and lengths, and limited helpers' storages with difference sizes
2. This challenging problem is NP-hard submodular function maximisation with multiple linear constraints
3. We have designed an efficient heuristic algorithm to allocate the buffer
   - Our algorithm achieves at least the same performance as the very high-complexity SFM algorithm, traditionally used to solve this type of challenging problems